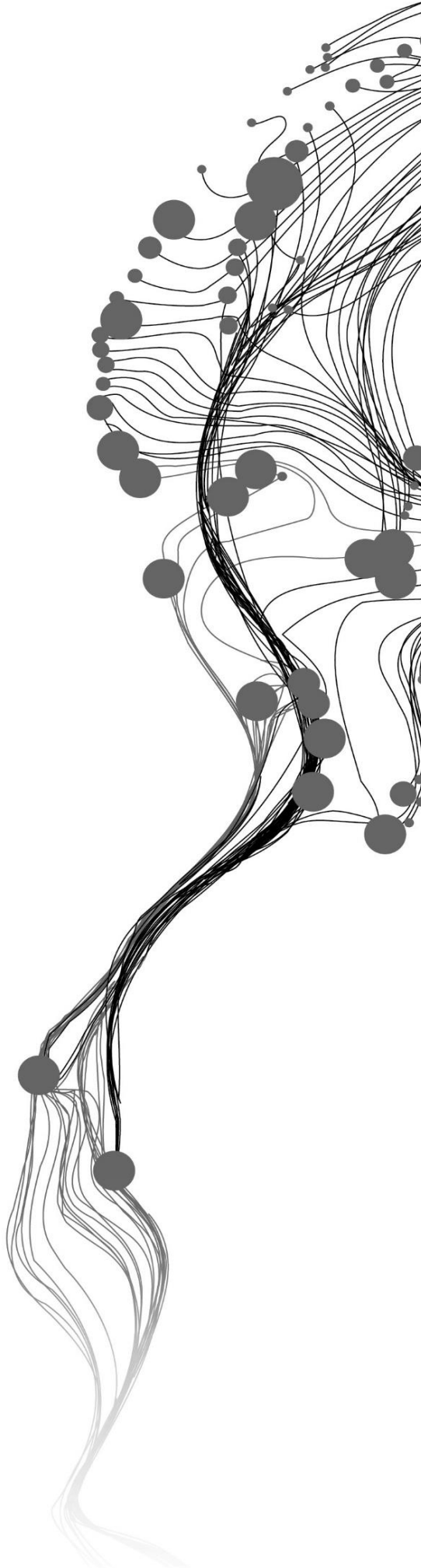


VISUAL ANALYTICS APPLIED TO CROWDSOURCED GEO- INFORMATION

JOTHIRNADH SAI GUTHULA
March, 2015

SUPERVISORS:
Prof. Dr. M. J. Kraak
Dr. Ir. R. L. G. Lemmens



VISUAL ANALYTICS APPLIED TO CROWDSOURCED GEO- INFORMATION

JOTHIRNADH SAI GUTHULA

Enschede, The Netherlands, March, 2015

Thesis submitted to the Faculty of Geo-Information Science and Earth
Observation of the University of Twente in partial fulfilment of the
requirements for the degree of Master of Science in Geo-information Science
and Earth Observation.

Specialization: Geoinformatics

SUPERVISORS:

Prof. Dr. M.J. Kraak

Dr. Ir. R.L.G. Lemmens

THESIS ASSESSMENT BOARD:

Dr. Ir. R.A. de By (chair)

Dr. C. Stasch (External Examiner, 52⁰ North)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

We are living in a world where the data collection is becoming ubiquitous and becomes part of day to day life, resulting in high volumes of data. One form of this data collection is crowdsourced geo-information that uses the Web as a resource to deal with the collaboration and sharing of geo related content. Ongoing research is putting an effort to differentiate this crowdsourced information in order to understand it and analyse it for further information extraction. The visual representation of crowdsourced data is not trivial because of its volume. Geovisual analytics plays a supportive role in understanding and analysing these large datasets. This research aims at identifying the Space-Time Cube for the visualization of crowdsourced geo data with an objective to evaluate the usefulness of that representation in discovering the spatio-temporal patterns. As part of the study, three different case studies on three crowdsourced applications (Flickr, enviroCar, and OpenStreetMap) are selected. The data required for the analysis of three case studies is collected and visualized in the Space-Time Cube (STC) for preliminary spatio-temporal analysis. For one case study (enviroCar) a usability analysis was conducted using the 'Questionnaire evaluation method' to judge the usefulness of the proposed visualization. The overall evaluation result showed that the cube had its advantages and disadvantages as well. It is very helpful for spatio-temporal analysis but, the disadvantages mentioned are due to the problems in understanding the visual complexity of STC.

Keywords:

Crowdsourced geo-information, Visual analytics, Spatio-temporal patterns, Space-Time Cube, Flickr, enviroCar, and OpenStreetMap, Usability analysis, Questionnaire evaluation method.

ACKNOWLEDGEMENTS

I would like to dedicate this work to Dr. P. S. Roy Sir who gave me a wonderful opportunity to enter ITC and my lovable parents who sacrificed their joy for my education.

First, I like to thank my supervisors Prof. Dr. M. J. Kraak and Dr. Ir. R. L. G. Lemmens for their kind help and support throughout my thesis. My special gratitude to MSc. Bass Retsios, for his support and dedication. Without his help this research would never be completed.

I would like to take this opportunity to thank all the people, whom I met and worked with during the past 18 months. Especially, I like to thank Luiz Andrade, Bernard (bro) and Jelena Marjanovic (Messy) for sharing their valuable time with me.

I would like to thank all my friends from India who are always there to cheer me up.

Last, but not the least, I would like to thank Ramya for being a part of me and sharing my good and hard times.

TABLE OF CONTENTS

1.	Introduction.....	1
1.1.	Motivation and problem statement.....	1
1.2.	Research identification	2
1.3.	Project set-up	4
1.4.	Thesis Outline	5
2.	Theoretical framework.....	7
2.1.	Introduction	7
2.2.	Crowdsourced geo-information.....	7
2.3.	Peuquet Triad Framework	9
2.4.	Present visualization methods	9
2.5.	Space-Time Cube.....	13
2.6.	Summary	15
3.	Case studies of this research	17
3.1.	Introduction	17
3.2.	Data consideration point of view	17
3.3.	Case study on Flickr data	17
3.4.	Case study on enviroCar data.....	19
3.5.	Case study on OpenStreetMap data.....	20
3.6.	Summary	21
4.	Visual representation applied to crowdsourced data	23
4.1.	Introduction	23
4.2.	Analyzing the crowdsourced data collected.....	23
4.3.	Visual Variables.....	23
4.4.	Application of visual variables to Space-Time Cube.....	25
4.5.	Summary	29
5.	Implementation of crowdsourced data	31
5.1.	Introduction	31
5.2.	Visualization environment used.....	31
5.3.	Implementing the collected crowdsourced data	32
5.4.	Summary	44
6.	Evaluation	45
6.1.	Introduction	45
6.2.	Selecting an evaluation method.....	45
6.3.	Usability evaluation on selected group and its objective	46
6.4.	Evaluation results	47
6.5.	Discussion of the evaluation results.....	54
6.6.	Summary	55
7.	Conclusion and further recommendation.....	57
7.1.	Introduction	57
7.2.	Conclusions	57
7.3.	Recommendations for further research.....	60
	Appendix 1: Questionnaire evaluation.....	63
	Appendix 2: Visualizations printed in A3.....	70
	Appendix 3: Code used to download enviroCar data.....	74
	Appendix 4: presentation Slides	76
	List of references	79
	URL	81

LIST OF FIGURES

Figure 1.1: Methodology.....	4
Figure 2.1: Crowdsourcing context.....	8
Figure 2.2: The fundamental components of the Triad framework.....	9
Figure 2.3: Minard's map of Napoleon's campaign to Moscow.....	10
Figure 2.4: OpenStreetMap.....	10
Figure 2.5: Static map representing the Sandy-related (Shelton et al., 2014)	11
Figure 2.6: Series of static maps representing the developments of Enschede	12
Figure 2.7: Space-Time Cube in ILWIS environment.....	14
Figure 3.1: Flickr interface.....	18
Figure 3.2: Representing enviroCar application.....	19
Figure 3.3: OSM data processing	20
Figure 4.1: Retinal variables proposed by Bertin.....	24
Figure 4.2: Size variable representation in STC environment	25
Figure 4.3: Shape variable representation in STC environment.....	26
Figure 4.4: Color variable representation in STC environment	27
Figure 4.5: Size variation in STC environment.....	27
Figure 4.6: Shape and color variation in STC environment	28
Figure 4.7: Shape variation of single track.....	28
Figure 5.1: Space Time-Cube interface in ILWIS environment.....	31
Figure 5.2: Graph representing the Flickr posts made by the users as per day	32
Figure 5.3: 2D view of Flickr posts	33
Figure 5.4: Interface to load data into Space-Time Cube.....	33
Figure 5.5: Implementation of Flickr case study in Space-Time Cube	34
Figure 5.6: Spatio temporal analysis on Flickr data.....	35
Figure 5.7: Graph representing the no. of posts made by the users.....	35
Figure 5.8: enviroCar data classification	36
Figure 5.9: enviroCar data in STC environment.....	36
Figure 5.10: enviroCar tracks classified on the basis of fuel consumption	37
Figure 5.11: Spatio-temporal analysis on enviroCar data	38
Figure 5.12: Identification of anomalies in STC.....	38
Figure 5.13: OSM data collected from different years in disaster effected regions	39
Figure 5.14: OSM data collected from different years in different regions.....	40
Figure 5.15: Legend of the OSM data.....	40
Figure 5.16: OSM contributed data of Miyagi Region in STC.....	41
Figure 5.17: OSM contributed data of Haiti region in STC	42
Figure 5.18: OSM contributed data of Overijssel in STC.....	43
Figure 5.19: OSM user contributed data of Tanzania in STC.....	44
Figure 6.1: Pie charts representing the evaluation results of part-1.....	48
Figure 6.2: Front view of STC.....	48
Figure 6.3: Finding anomalies in STC	49
Figure 6.4: Spatio-Temporal Analysis in STC.....	50
Figure 6.5: User answer on the identification of anomalies using STC.....	50
Figure 6.6: User answer to the identification of spatio-temporal patterns in STC	51

Figure 6.7: Flickr data analysis.....52

Figure 6.8: Answer to the identification of spatio-temporal patterns of Flickr data using STC53

LIST OF TABLES

Table 4.1: Measurement Scales (Kraak & Ormeling, 2011)..... 23

Table 4.2: Organization levels of Graphic variables 24

1. INTRODUCTION

1.1. Motivation and problem statement

We are living in a world where data collected from our daily life are now practically ubiquitous and easily available from diverse data providers, mostly through the web. Due to the recent advancement in the field of GEO-IT and easiness in the usage of data creation tools with Web 2.0 services, users are willing to create, contribute and collect geographical information from their daily life activities which have social impact, with the help of crowdsourcing platforms (Díaz, Granell, Huerta, & Gould, 2012). Most of the volunteered data contributed by citizens is available through the web. Web portals such as Google Map Maker, Wikimapia and OpenStreetMap (OSM) are encouraging volunteers to create global patchwork of geographic information. Organizations like enviroCar.org, Tekenradar.nl are trying to gather crowd sourced data for global environmental assessments by naming the volunteered individuals as the citizen scientists. Considering these factors Goodchild for the first time gave a name to this phenomenon as Volunteered Geographic Information (VGI) (Goodchild, 2007a, 2007b). He did an analysis on associated issues like, what drives people to do this? What is the completeness of this data? How is this going to threaten the privacy of the individual? And how can VGI enhance more conventional sources?

Volunteers collect most of this data, and this crowdsourced data is available through the web. However, the quality of this data is still a problem as it has no quality control, no metadata, and no standards to be followed. Even though Goodchild and other researchers provided particular solutions for this problem (Elwood, Goodchild, & Sui, 2012; Goodchild, 2012; Heipke, 2010), it needs further research for its justification and analysis. As this data is growing exponentially in everyday life, researchers and private sectors have begun to focus on how to utilize this data in understanding the potential insights of spatio-temporal patterns. But the data available in this way is not easy to analyze for information extraction because of its abundance. For this reason, how to visualize these large datasets for easiness in the assessment of spatio-temporal data became an essential mission. Since, the data collected by crowdsourcing is so diverse, so getting insight into this data is problematic. In this case, recent developments in visualization, such as visual analytics helps in finding a solution.

Visual analytics is a science that “combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets” (Keim, Kohlhammer, Ellis, & Mansmann, 2010). This term was first used in the book ‘Illuminating the Path’ (Thomas & Cook, 2005), even though the ideas and research approaches emerged much earlier. According to Thomas, visual analytics is a multidisciplinary field which includes analytical reasoning techniques, interaction techniques, visual representations, data representations, transformations and the methods to support production, distribution, and presentation of the results. This concept spread to other disciplines of research like information visualization and geo visualization.

With the advent of Geographic Information Science (GI Science) which involves computational processing of large datasets for graphical representation, requires a field like geovisual analytics for increasing the speed of the data processing and also for analysing large datasets (Andrienko et al., 2010). Geovisual analytics (GVA) is a part of visual analytics that is used in viewing this large dataset in a different perspective for spatio-temporal pattern analysis. GVA is be used for identifying the known and discovering the unknown with understandable assessments (Keim et al., 2010). GVA makes use of many existing visual representation and integrates them with spatio-temporal analysis like the Space-Time cube (STC).

The Space-Time Cube (STC) is a visual representation that can be a part of a GVA environment in which any type of visualization can be a part and can be used in visual spatio-temporal pattern analysis.

Hägerstrand (1970) introduced the STC for representing both spatial and temporal components of human activities at a time, where the horizontal axis represents the spatial component, and the vertical axis represents the time component. The STC is mainly used to visualize movement data (Kveladze, Kraak, & van Elzakker, 2013). It is also possible to visualize and explore the aggregation of movements with the use of an STC, where trajectories are shown as 3D polylines through space and time (Demšar & Verrantaus, 2010). The concept of 3D space-time density of trajectories is used for this analysis. So far the crowdsourced data has not been linked too much to space and time, so let's look back into the GVA visualizations where opportunities to bring them together can be seen.

Research is being done in visualizing the crowdsourced data using different visualization methods. Researchers are trying to use the social media data or the crowdsourced geo data for understanding the spatio-temporal patterns according to their analysis. But the main question arises, how successful they are in utilizing this data, how far these visualization techniques are exactly useful in representing the spatio-temporal patterns. Most of these researchers visualized this crowdsourced data using different visualization methods like static maps, series of static maps, and animated maps. But all these visualization methods are limited to either spatial component or temporal component or in some cases both of these components but represented using different map panels at a time. Using a series of static maps one can represent temporal aspect to some extent, but it is limited to the number of images in use. Animated maps are very helpful in representing the spatial and temporal aspects, but it is limited to some aspects like; change of scale in maps and comparison of different temporal maps. Tsou & Leitner, (2013) in their research mentioned that, some reviewers and researchers realized that traditional mapping methods and social analysis tools might be challenging for mapping social media content as they have “unique spatial distribution patterns, dynamic user-generated locational information, and complicated social networks”.

In this context, considering the problems and suggestions given by previous researchers and taking into account the advantages of the STC, this research is motivated towards experimenting with the STC to represent spatio-temporal components of different crowdsourced applications. These applications are selected on the basis of three different views on the generation of the crowdsourced content, where some questions on the applications like, why this application is selected specifically? What is the purpose of generating this application data? How useful is it? Will be explained in Case Studies chapter (Chapter- 3). This can lead to more conventional ways of understanding the crowdsourced data.

To realize the work in this research, three different case studies are selected from three different crowdsourcing applications. These three crowdsourcing applications are Flickr, enviroCar, and OpenStreetMap that were discussed in detail in the coming chapters. The data collected through these applications were visualized in the Space-Time-Cube for understanding the spatio-temporal analysis.

1.2. Research identification

1.2.1. Research objectives

The main objective of this research is to evaluate the usefulness of Space-Time cube (STC) to discover the spatio-temporal patterns in the context of different crowdsourced applications.

This objective can be reached by defining the following sub-objectives:

- a) To identify and describe the character of different types of crowdsourced data along with the aspects of completeness and trustworthiness.
- b) Discuss the proper visual representations that can deal with time and space, with a focus on the STC.
- c) To elaborate on three case studies that can be used for pattern discovery using the selected visual representations.
- d) To evaluate the usefulness of the enviroCar case study, if the visual solutions offered in this research are understood by the users.

1.2.2. Research questions

Related to sub-objective a:

1. What are the different types of crowdsourced geo-information?
2. How is crowdsourced data produced, and what are the consequences for its quality?

Related to sub-objective b:

3. What are the most commonly used visualization techniques for the crowdsourced data analysis?
4. Among those visualizations, which are best suited to deal with the character (location, attribute, and time) of crowdsourced data?
5. How does the STC fit-in?

Related to sub-objective c:

6. What is the objective of the enviroCar case study, and what is the nature of its data?
7. What is the objective of the Flickr case study, and what is the nature of its data?
8. What is the objective of the OSM case study, and what is the nature of its data?
9. How can the data of each case study be implemented into the STC?

Related to sub-objective d:

10. How to evaluate the results?
11. How to execute a questionnaire session?
12. How was the solution experienced by the user of the case study?

1.2.3. Innovation aimed at

The innovation of this thesis topic is aimed at finding the added value of using the Space-Time Cube for spatio-temporal analysis of crowdsourced geo-information and to differentiate between cases of crowdsourcing in which STC can be worth using.

1.3. Project set-up

1.3.1. Method adopted

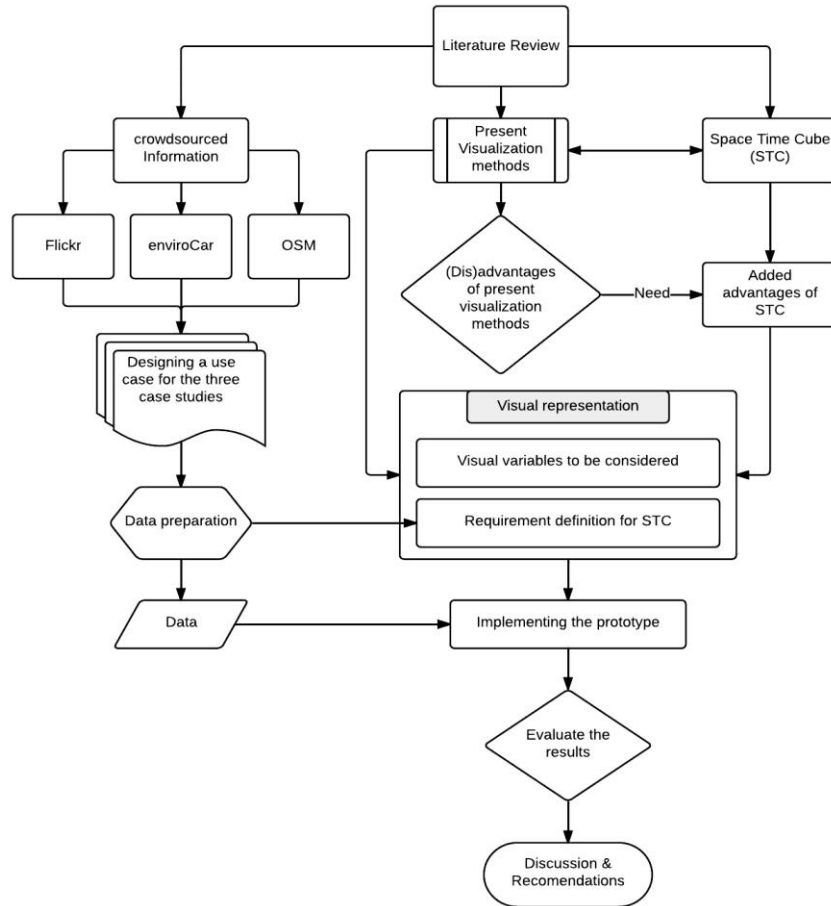


Figure 1.1: Methodology

The following phases of activities (see Figure 1.1) are executed to meet the present objectives of the research:

Literature Review:

Literature related to the concepts of crowdsourced geo-information, the Citizen Science, the Volunteered Geographic Information (VGI), the existing visualization methods (Single static maps, series of static maps and animation maps), different GVA visualization methods in use like the Space Time Cube (STC), the methods for implementing visualization techniques, crowdsourced applications like Flickr, OpenStreetMap and enviroCar will be reviewed and discussed. The case studies will be drawn from the literature review. The literature review will contribute to this research in forming the theoretical background. Depending on this phase, the main disadvantages of the traditional methods available will be identified. Also the applications of the STC and its usability will be identified. One of the most important steps in this phase is studying different methods of conducting usability analysis in order to evaluate the model. This phase will act as a part throughout all the phases of research.

Data preparation:

Since most of the crowdsourced data is available through the web, any individual can download it at any time. This research is going to concentrate on three case studies that deal with the enviroCar data, which can be downloaded from enviroCar API (“enviroCar,” 2013), Flickr posts, collected with the help of Flickr API (Ludicorp, 2004) and OpenStreetMap data, downloaded from OSM database (“OpenStreetMap,” 2004). The data will be downloaded according to the purpose of the research as the data is abundant, and

the data will be limited according to the need of the research. Considering that this is volunteered crowdsourced geo information one cannot rely on it completely. For this reason, the datasets are needed to be analyzed for errors and cleaned in order to prepare it according to the conditions that crowdsourced data must comply to be used in the STC. This phase will deal with the research questions 6, 7, 8, and 9.

Conceptual framework:

In this phase, a conceptual framework of the visualization methods to be applied within the STC for analysing the spatio-temporal patterns of the crowdsourced data will be decided. From the literature review the mostly used visualization techniques in the STC will be identified and try to implement them in this research depending on the different type of information. The requirements of the STC must be identified to visualize the available data due its visual complexity. For this reason the visual variables must be selected that are best useful for visualization. In this phase the research questions 3, 4, and 5 will be answered.

Implementation of crowdsourced data:

In this phase the crowdsourced data is imported into the ILWIS software. Once the data is imported, the 3D option and the STC option will be checked to implement the data. The different visual variables described in the previous phase will be executed on the crowdsourced data for spatio-temporal analysis. Depending on the nature of the data different variables are used for different datasets. Based on the spatio-temporal results obtained, the most appropriate representation will be selected for final analysis. OpenStreetMap will be used as the basemap for this analysis. This phase will provide answers for the research questions 5 and 9.

Evaluating the spatio-temporal results:

Once the implementation part is completed, the evaluation phase will be started. The knowledge gathered from the literature review will be utilized on usability analysis and an evaluation method will be selected. The usability analysis will be executed on a group of persons with the help of selected evaluation method. The group will be selected on the basis of the expert knowledge on one case study. Depending on their availability of participants the evaluation will be executed and the results will be summarized. This phase will answer the research questions 10, 11, and 12.

Discussion and Recommendations for further research:

After completing the above phases, depending on the results obtained from usability analysis a research conclusion can be derived. Usefulness of the STC will be evaluated depending on the results obtained from the usability analysis. In this phase also the problems encountered at the time of analysis will be discussed. The recommendations for further research will be summarized at the end.

1.4. Thesis Outline

Including the introduction, this thesis is organized in seven chapters. Chapter 2 discusses about the complete theoretical background of this research. This chapter serves two purposes, first to identify the different types of crowdsourced applications and second to judge the usefulness of STC compared to other visualization methods. In Chapter 3, the triad framework will be introduced, and the structure of the three case studies will be discussed. Chapter 4 is about understanding the cartographic information, present visual variables and their applications on Space-Time Cube for the better visualization of crowdsourced data. Chapter 5 is about implementing the selected crowdsourced data into the STC. In this chapter a preliminary analysis is conducted on the implemented data for evaluating the representation. Chapter 6 discusses about the selected evaluation method and the results achieved from that evaluation analysis. Finally, in Chapter 7 the conclusion and the recommendations for further research are discussed.

2. THEORETICAL FRAMEWORK

2.1. Introduction

The theoretical framework focuses on the explanation of different concepts and the background information required for understanding this research. In this chapter concepts like crowdsourcing (types and quality aspects) and the present visualization methods will be described. Then the Time Geography, the Space-Time Cube (STC) and the Triad framework will be introduced. In the following chapter, three case studies will be considered for evaluating the usefulness of the STC. After that a chapter on the graphical representation will be described, in which position, size, value, color, orientation, and texture are used for the representation of these case studies in STC.

2.2. Crowdsourced geo-information

Crowdsourcing has multiple interpretations, but it can be termed as a process that deals with the collaboration and sharing of information through the participation of large groups of people, using the web as a resource. Howe (2008) stated crowdsourcing as the term often used in data creation methods, where the data is generated and shared by the large groups of users who are not prearranged. With the development of the web 2.0 technologies, the sharing of data became more prominent through the social networking sites like YouTube, Twitter, Facebook, and communal databases like Flickr. These technological developments gave a direct push on crowdsourcing with the designers or creators or the managers of these sites (Hudson-Smith, Batty, Crooks, & Milton, 2009) to pull the crowd. While telling the story of crowdsourcing it ends up with many applications that are beyond the scope of this research, so we are limiting our study only to the geo information gathered through crowdsourcing.

Crowdsourced geo-information is a part of crowdsourcing that uses the web as a resource to deal with the collaboration and sharing of geo related content. Ramm (2009) gave his idea on how crowdsourced geo information evolved into crowdsourcing by taking OpenStreetMap as an example. Again with the advancements in the field of Geo-IT and the web 2.0 technologies, researchers came up with different terms and definitions for crowdsourced geo information. Among those terms, the most leading phenomena are “Citizen Science and Volunteered Geographic Information (VGI)”. These phenomena are revolutionizing the map creation and their usage (Goodchild, 2007a) in the recent era. Heipke (2010) states that the main differences are the fact that the volunteers, deficient of formal training in map creation, produce the geospatial data themselves rather than depending on the commercial services. Because of this reason change detection occurs in real time as large user group’s work together voluntarily for mapping. Ramm (2009) conferred how crowdsourcing and mainly crowdsourced geo information had revolutionized since 2005 by taking OpenStreetMap as an instance.

As one would expect for new fields, like Volunteered Geographic Information and Citizen Science have many different definitions. Citizen Science can be defined as “Public participation in organized research efforts” (Dickinson, Shirk, & Bonter, 2012). While in OpenScientist blog, Citizen Science is described as “The systematic collection and analysis of data; development of technology; testing of natural phenomena; and the dissemination of these activities by researchers on a primarily avocational basis” (OpenScientist, 2011). While in Muki Haklay’s personal blog, Haklay (2007) noted that “[Citizen Science is defined as] ...scientific activities in which non-professional scientists volunteer to participate in data collection, analysis and dissemination of a scientific project...”. In 2007, Goodchild originated a new discipline for user-generated content through general Web phenomenon and termed it as Volunteered Geographic Information (VGI). According to him the VGI is the engagement of large groups of volunteered citizens,

lacking any technical training or qualifications, in the creation and generation of geospatial information (Goodchild, 2007a, 2007b). The results generated through this process may or may not be accurate, but their average can be matched by the expert judgment (Hudson-Smith et al., 2009).

It is thus clear that both disciplines are termed depending on the phenomenon of content generation by large number of citizens. To summarize: In citizen science the volunteers collecting data, have a particular goal to achieve and they also have some fundamental knowledge through which they can meet their needs, but with respect to VGI, the volunteers may or may not have any goal to collect the volunteered geographic information; it is not necessary to have any prior knowledge of this data collection. The different terms of crowdsourcing and their differences are illustrated in figure 2.1.

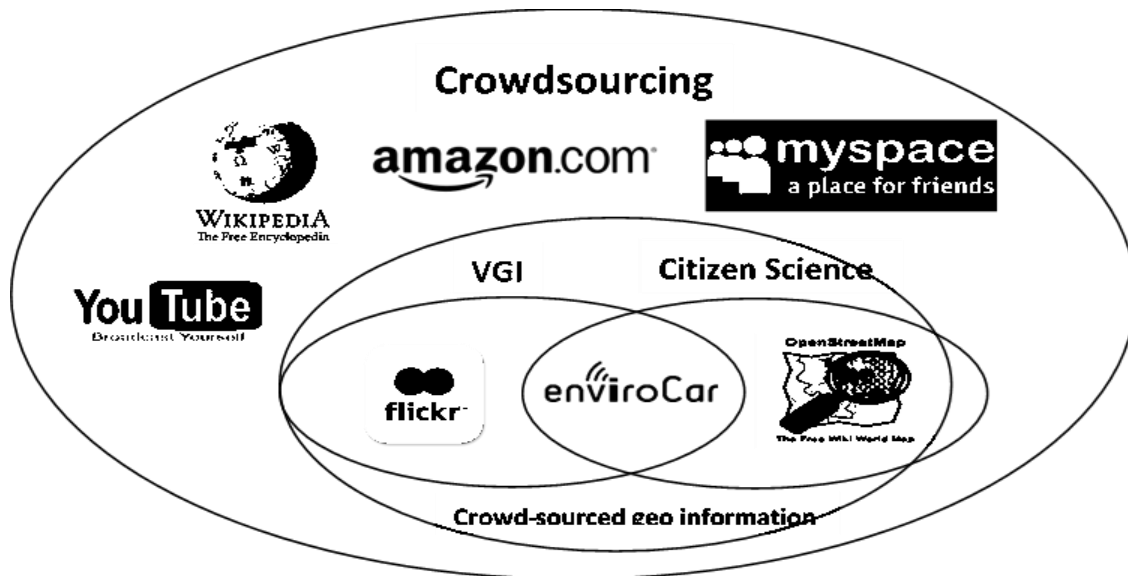


Figure 2.1: Crowdsourcing context

In the upcoming chapters, these terms will be further explained by considering three case studies on Flickr, enviroCar and OpenStreetMap. Nielsen (2006) observed that 90% of the social media users just consume the information; 9% users contribute information occasionally, and 1% users participate a lot and contribute a lot of information as they don't have lives. He also named it as "90-9-1 rule for participation inequality in Online Communities and Social Media" at his time. Considering the user participation to this extent reminds us of the quality aspects of the data.

The quality of the crowdsourced data is an important factor to consider, as the data is produced by large group of users, who might or might not have any prior knowledge of the application. This factor might lead to a lot of quality issues, but the term quality is described according to the "fitness for use" of the application (Heipke, 2010). Most of the crowdsourced geo information created through GPS receivers produces GPS tracks and orthophotos (Goodchild, 2007a) which are mostly accurate. Local knowledge also plays a significant role in quality aspects of the data as most of the crowdsourcers are interested in mapping their surroundings first (Heipke, 2010), which makes the data quite accurate. Even if the individual estimates of large user groups are incorrect, their average can be considered to be accurate, which is the main advantage of crowdsourcing. Also, Surowiecki (2004) tried to select randomly some estimations, collected from large user groups lead to the data that is surprisingly accurate. Though there is no perfect method for calculating the quality of crowdsourced geo information, some researchers tried to find the quality depending on the purpose of different crowdsourced applications.

The crowdsourced information has the time factor describing the temporal information, the spatial factor representing the spatial extent of the information and the attributing factor explaining who or why did he/she gather the information. For understanding, this kind of spatio-temporal data Peuquet (1994) introduced a triad framework, which is discussed in the following section.

2.3. Peuquet Triad Framework

Peuquet (1994) proposed the Triad framework with a goal of representing spatio-temporal data in a way that conforms human conceptualizations on an attribute perspective. According to Peuquet, the framework contains three fundamental components of data: Time (when), Location (where), Attribute (what). Crowdsourced data can be captured by the triad framework for understanding the spatio-temporal patterns as the data has the spatial (where), temporal (when) and attribute (what) information. The three basic questions formed with the help of the Triad frame can be used in understanding the data (Peuquet, 1994).

- **When + where → what:** Describe the objects or set of objects (what) that are present at a given location or set of locations (where) at a particular time or set of times (when)?
- **When + what → where:** Describe the location or set of locations (where) occupied by a given object or set of objects (what) at a particular time or set of times (when)?
- **Where + what → when:** Describe the time or set of times (when) that a given object or set of objects (what) occupied a particular location or set of locations (where)?

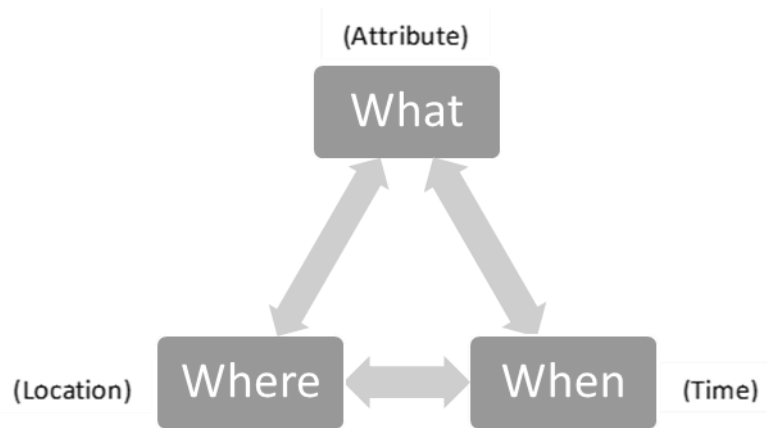


Figure 2.2: The fundamental components of the Triad framework

Figure 2.2 represents the fundamental components of the Triad framework. As the triad framework systematically helps us to formulate the basic questions that are applied to the three case studies for understanding the nature of the crowdsourced geo information will be discussed in the upcoming chapters. In the following section, the present visualization methods will be described in detail with the help of some examples and followed by introducing the concept of time geography into crowdsourcing.

2.4. Present visualization methods

Depending on the task, every visual representation has its advantages and disadvantages on how the representation supports certain degree of explorations (Kraak & Madzudzo, 2007). The usefulness of the visual representation depends on the fitness for use, so before applying any visual representation it is important to note the advantages and disadvantages of such representation. Traditionally in the past era, researchers only used to visualize the information using the paper maps. But with the advancement in the field of GIS and the development of new software's helped the users to visualize the information using different visualization methods like series of static maps and animation maps along with the single static maps.

(1869) on the Napoleon's doomed campaign to Moscow during the period of 1812-1813. Tufte considered this map as one of the classic static maps ever (Kraak & Ormeling, 2011). In this map, Minard visualized the space, the time and the attributes like temperature and death of the soldiers simply in a two-coloured map.

In most of the crowdsourced information visualizations, users use different geovisualization techniques to extract the information. For example, they use the graphs to represent the temporal patterns and single static maps to represent the spatial information. The combination of these two leads to the identification of spatio-temporal patterns (Neis & Zipf, 2012). Shelton, Poorthuis, Graham, & Zook (2014) tried to visualize the Sandy related tweets made in the United States region at the time of the disaster (Figure 2.5). The complete spatio-temporal data of this disaster (tweets from 24th October, 2012 to 31st October 2012) is represented in single static map. The polygons on the map represent the intensity of the tweets made in that region.

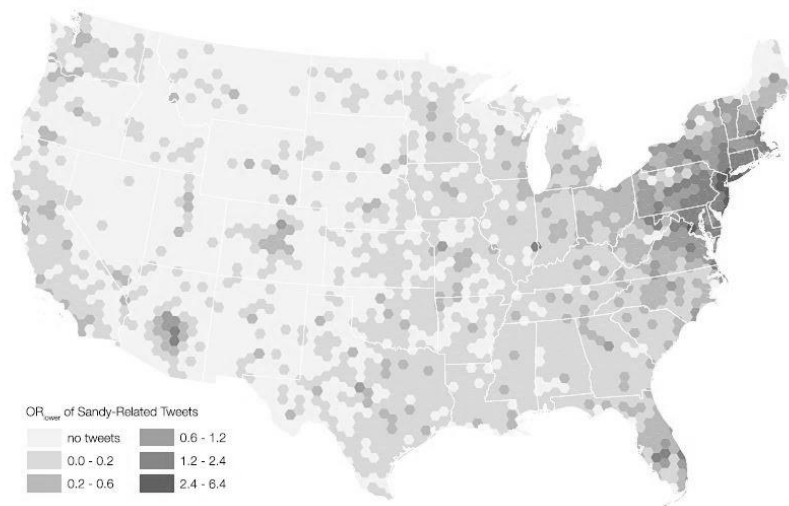


Figure 2.5: Static map representing the Sandy-related (Shelton et al., 2014)

Today, GIS software's helped the users by making the process of preparing a static map easier for presenting the geospatial information, through which the traditional role of a map as a presentation device is being preserved. But, according to Kraak (2003) a map should act as an interface that can handle the information access and exploratory activities instead of being a presentation device. Because of this reason users started to search for new visualization methods that can deal with the exploration of the data and information accessing.

The crowdsourced data collected through the GPS consists of attribute information along with the time and location information. So a standard static map can be used as a base map on which the time and attribute information will be visualized, which makes the key issues for visualization.

2.4.2. Series of Static Maps

A series of static maps is the combination of different static maps where every individual static map represent images of different time (Kraak & Ormeling, 2011). This combination of static maps is used to describe an event. While comparing these individual maps, the change in space according to time is analyzed. The user must follow the temporal sequence, which is represented by spatial sequence in order to distinguish the temporal changes. Figure 2.6 describes the developments (City expansion) of Enschede city for different time periods.

The number of images in the series of static maps is limited as it is hard to deal with the long series. And also in the past the data is limited, so the regular changes are not available to represent in continuous order. As a result, it can be noticed that the changes in the city cannot be represented using the series of static maps. Even if it is represented, there used to be a lot of temporal difference between one static map to the other. Although the complete change cannot be represented by the series of static maps, it represents the exact changes in time with specified snapshots that are helpful for the users in comparing them.

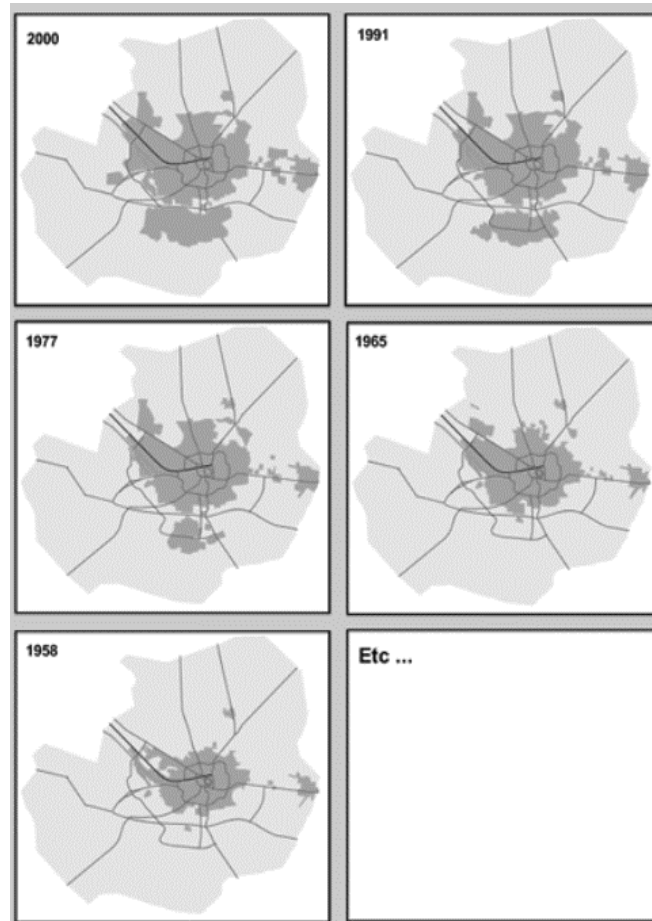


Figure 2.6: Series of static maps representing the developments of Enschede

2.4.3. Animated Maps

An animation map displays a series of static maps one after the other in a single frame, where every static map represents change via snapshots of time. This change might be a temporal change or a non-temporal change (Kraak & Ormeling, 2011). Cartography has paid its attention to the field of animation since the 1960s in the form of cartoon-like approach. With the advancement of technology in 1980s, the cartographic animations were developed from computer based imagery.

The animation maps are classified into two types, namely temporal animations and non-temporal animations (Kraak & Ormeling, 2011). A temporal animation represents the change in space with respect to time, like the changes in Enschede city from 2000 to 2014, whereas non-temporal animations represents the changes in space or attributes like the animation representing the alternative classes of a thematic map or fly-through. This research concentrates on the spatio-temporal data, so temporal animation is our primary concern.

Animation can be a better option for understanding specific applications like disasters. It can highlight the place of occurrence of the disaster and continue with the changes occurred in the surrounding regions in time. This representation gives a clear overview on the disaster, but it would be hard to interpret

each and every change occurring in that situation. The animations (URL 1, 2, 3) developed for representing the OpenStreetMap contributions on the Haiti region at the time of Haiti earthquake are perfect examples. Hence, the 2D animations cannot be helpful to answer all the questions of the users in complex situations.

Static maps, series of static maps and animation maps help the users in understanding the spatial features, as well as temporal features to some extent. However, there are limitations in all these representation methods which are making the users, to develop new representation methods in order to overcome these limitations. It would be helpful if there exists a way that can combine the spatial and temporal factors at the same place and also allowing the user to conduct different analysis on these factors. Space-Time Cube is one of the well-known spatio-temporal representation method which can give the answer to such questions. In the past, Space-Time Cube is mostly used to represent the movement data, but in this research, we are going to use it to represent crowdsourced geo information. In the following sections, the detailed description on the Space-Time Cube and the time geography were discussed.

2.5. Space-Time Cube

Before talking about Space-Time Cube (STC), it is important to know about the Time geography. Time geography was introduced by Hägerstrand (1970) at the end of sixties for understanding the social interaction and movement of individuals in day to day activities. Time geography justifies movement as a universal behavior of individuals. The patterns extracted and analyzed from the day to day activities of individuals might give rise to the understanding of other behaviors. Before the implementation of time geography, both spatial and temporal factors are considered separately for analysis, but time geography sees both time and space as inseparable. Now a day's time geography is being applied in many fields associated with transportation, anthropology, time user research, environmental science, geography and regional planning.

Introduced Space-Time model as the most prominent visualization model for the representation of time geography; where, it provides an efficient visualization environment for the individual movement patterns. In his three-dimensional Space-Time model, the individual movement is represented by a life path which is, "starting at the point of birth and ending at the point of death", by collapsing the three-dimensional space into a two-dimensional plane and representing time in perpendicular direction. Though, both space and time have their limits in the representation of individuals; it might be impossible to offer a broad classification of constraints. But, Hägerstrand grouped most of these constraints into three categories, namely: "the capability constraints" (for instance, the distance travelled or the accessible range), "the coupling constraints" (for the comparison of two entities) and "the authority constraints" (indicates the entities with both space and time) (Hägerstrand, 1970). Each constraint is well defined and has its own authority.

The Space-Time model also called as Space-Time Cube (STC) is a visual representation used by Hägerstrand in time geography. Fundamentally it looks like a cube where, the cube's horizontal plane (x-axis & y-axis) represents the geographic region (space) and the 3D vertical axis (z-axis) represents the time (Hägerstrand, 1970; Kraak, 2003). The main concepts related to the STC are:

- **Space-Time Path (STP)**- Represents the trajectory of the movement
- **Space-Time Prism (STPrism)**- Represents the areas to be reached within a time interval
- **Potential Path Space (PPS)**- Represents the radius covered by the STPrism
- **Potential Path Area (PPA)**- Represents the footprints PPS on the horizontal base map

STC provides a better environment to represent spatio-temporal data. But during the sixties, when the Space-Time environment was introduced by Hägerstrand the graphical options were limited, and the draftsperson used to do operations manually. Later with the advent of GPS and GIS software's the data collection, and its implementation had become easier (Kraak, 2003). Kraak (2003) extended the dynamic, interactive STC environment by including the options like full flexibility in viewing the STC, querying the data, moving the slider planes along the axis and highlighting the time period. Figure 2.7 shows the fundamental representation of the STC in the ILWIS environment.

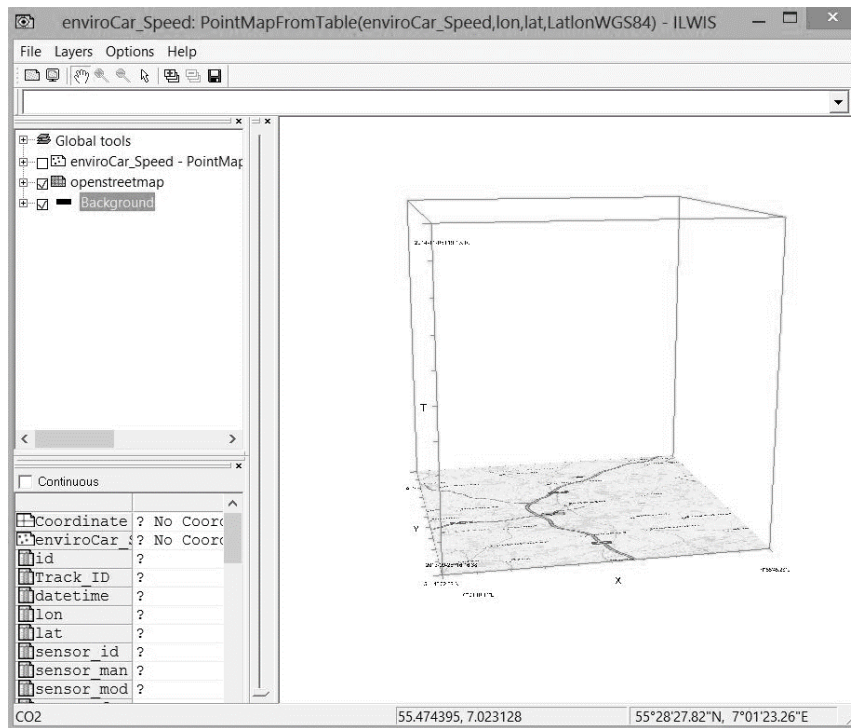


Figure 2.7: Space-Time Cube in ILWIS environment

Although the application of time geography in regional science started from 1970, there is an increased use of STC in the last decade. Miller, (1999) applied the constraints-oriented approach, using the concept of Space-Time Prism for finding the accessibility in transportation and planning. Kwan (2000) used this method for understanding the human activity-travel patterns in spatiotemporal dimensions for different genders and ethnic groups in Portland metropolitan region. Kraak (2003) extended the Space-Time Cube from the geovisualization perspective by including the options like data querying, option to move the slider, highlighting the required time period and implementing the cube in different windows increased the usefulness of STC. Kraak also demonstrated his results for various applications based on the latest developments in geovisualization. Kraak & Madzudzo (2007) suggested an alternative way to visualize epidemiological studies with the help of STC in their research. For this research, they took the Black Death case study for displaying the usability of the STC in health studies and compared their approach to present traditional visual representations. Kveladze et al., (2013) noted that there is very few research done to know whether the cube is truly operative and proficient to display complex datasets. So they designed a methodological frame work to evaluate the STC using the usability analysis. This frame work is used to evaluate the design aspects of the STC and also about the environment on which the STC is going to work with the help of four different case studies.

The main advantage of STC, when compared to the rest of the visualization methods, is its dynamic environment, which represents the patterns in a temporal dimension. The user can visualize all the data at a time in the cube and then conduct the analysis on the data by changing different base maps. They can also

compare different data sets, or choose different time periods and locations. STC, in most of the cases, is used to visualize trajectory data, where the path in the cube represents the individual movement. While in this research, the attempt to visualize crowdsourced data into STC where no simple path will be formed might give rise to a new utility approach. Also, the essential components of the spatio-temporal data in STC could provide what (attribute) and where (location) information by including the time (when) in order to answer data related questions. This analysis can be achieved by applying the Peuquet triad framework approach in STC.

2.6. Summary

In this chapter, the crowdsourced geo information, and the present visual representation methods were introduced and compared. There are many visual representation methods to deal with the spatio-temporal data, but all these representations had their limitations. Finding these limitations of other representation methods, in this chapter we reached a conclusion that considering the Space-Time Cube to represent the crowdsourced geo information is an interesting option for the object of study in this research. Therefore, STC is used as the visualization method in the following chapters. However, the visual variables used to represent the visualization and the conceptual framework to be applied in STC environment will be discussed after describing the case studies chapter.

3. CASE STUDIES OF THIS RESEARCH

3.1. Introduction

In the previous chapters, the nature of data, the time geography and the usefulness of the STC were introduced, but the question is how we can integrate VGI data with in STC environment. In order to understand that, in this chapter we will consider and discusses three different case studies of different types of VGI data to see how VGI data fits into STC. The Triad framework, the purpose of considering this case studies, nature of the data will also be discussed in this chapter.

3.2. Data consideration point of view

The complete description of the crowd sourced geo-information, citizen science, VGI and its different types were discussed in the previous chapters. According to the type of crowd sourcing, the nature of the geo-information will differ. Considering this as the major point of view, in this chapter we considered three case studies on different kinds of crowd sourcing application. The three case studies considered are: Flickr, enviroCar and OpenStreetMap. All three case studies comes under the crowdsourced applications, but on additional study they can be further distinguished.

Flickr (Ludicorp, 2004) is a crowdsourced application in which users can share their photos or videos with other users without any specific motive or intention. The photos shared will be having tags, comments and notes. The tags given for the photos will be used in organising the data and for scientific analysis. For scientific analysis the participants sharing the photos does not require any technical training or previous knowledge. Certainly the users don't have any idea on how this information is going to be used after. Because of this reason Flickr comes under the VGI applications according to the previous definition of VGI.

enviroCar (2013) is a crowdsourced application for citizen science projects, in which volunteers are collecting the data to get insight into sustainable mobility. This collected data can be used by citizen scientists for scientific analysis. Here the volunteers need to have some instruments to collect the data which will be explained clearly in the coming context. Here the volunteers collecting the data might not have the complete knowledge on the analysis to be done later. Considering the necessity of the instruments for collecting the data and the level of knowledge of the participant on the scientific analysis to be done on the data, enviroCar comes at the intersection point of VGI and citizen science projects.

OpenStreetMap (2004) is a crowdsourced project in which volunteers are participating and contributing their data with a specific objective of creating a free world map. This contributed data from all the users is helping in designing a world map from scratch since 2006. For participating in this project the volunteers need to have some knowledge on the map digitization part and the user might also need basic equipment like GPS for the data collection. Even though OSM is a crowdsourced project it needs the knowledge of the application and equipment for the data creation and contribution. Because of this reason OSM can be classified as a citizen science project according to the previous definitions.

These three case studies are a part of the crowdsourced geo-information starting from VGI with the help of Flickr, showing the intersection part of VGI and citizen science projects with the help of enviroCar and continued to citizen science project with the help of OSM maps. All three case studies will be explained in detail with the help of Peuquet triad framework in the coming context.

3.3. Case study on Flickr data

Social networks are defined as "A set of people who can share a common interest and have connections of some kind, among interacting units" (Wasserman & Faust, 1995). Generally the social media data consists of four attributes, they are: profile, friends, their relation and posts. These four attributes

represent the same characteristics in most of the social networks but were named in different ways, like “posts” in Facebook and “tweets” in Twitter. The geographic content in the social media has increased with the advent of location-acquisition technique in the recent era (MacEacheren et al., 2011). This lead to the huge development in gathering VGI data through social media sites. In this case study geo-located posts of Flickr at the time of Sandy Hurricane disaster are selected as one of the social media data for analysis in STC.

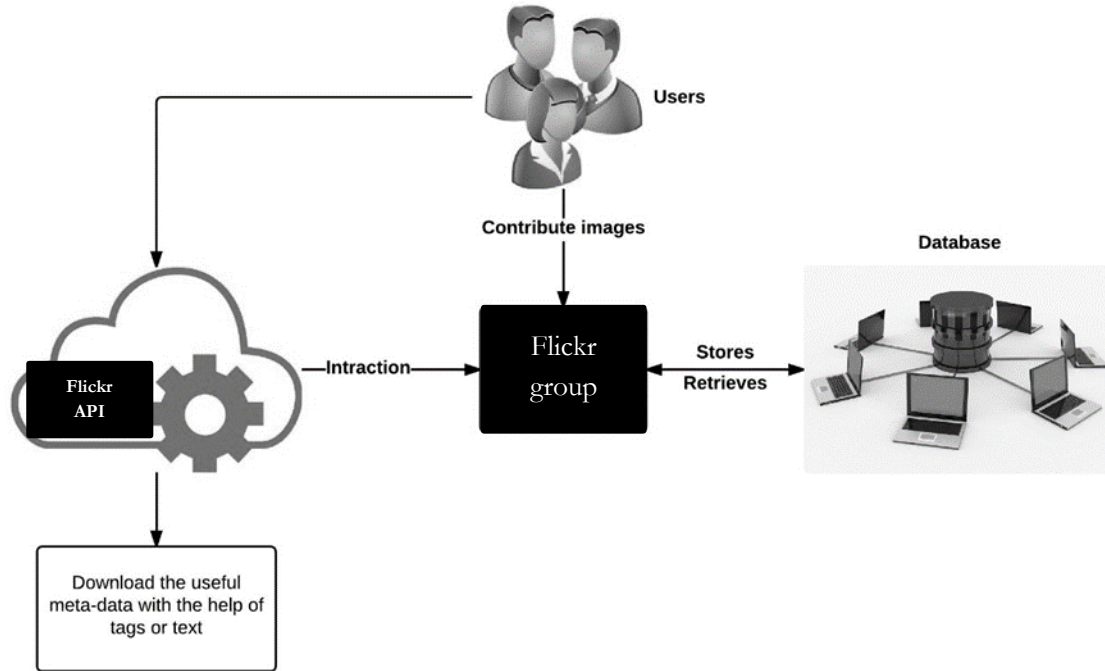


Figure 3.1: Flickr interface

Flickr (Ludicorp, 2004) is a social media application, which is being used for studying socio economic patterns of human behaviour in the recent years. This might be because of the global scale usage of Flickr by the users. This social media application is not created to collect geo information, but researchers generally extract the posts related to specific field of interest to do their analysis, hence the data available through this application is not a direct form of VGI but the researchers can extract the VGI from it. The above figure 3.1 represents the Flickr interface. In Flickr the users will contribute the pictures by posting them in their profile. The pictures posted might contain some hashtags and text description. These pictures are stored in a database from which the users can access them back at any time. Every user is allocated with one terabyte of personal space to store their pictures and videos. Flickr also designed an API for the users to retrieve the data from the server (URL 4). Flickr interface is represented in figure 3.1. The data to be downloaded can be sorted out with the help of Hashtags and text posted along with the pictures.

In this case study the Hurricane Sandy disaster situation which happened in United States in 2012 is selected. The reason for selecting this disaster as one of the case study is because Sandy Hurricane disaster is one of the most used disasters in social media in the recent years and also to understand how the social media is used by the users at the time of disaster. Considering this disaster might give us an insight on how Flickr can be used in understanding the human behaviour at the time of disaster. The data analysed in this case study is restricted only to geo-tagged posts in order to avoid problems on data quality, precision of the post, geocoding accuracy and to avoid the delusion between a user’s home place and the position from where a particular post is posted. In this case study it is considered whether by harvesting space and time data generated from the posts of Flickr in the STC along with the Hurricane movement data might contribute to a better understanding of the disaster.

Since only the geo-tagged posts are considered in this case study, the data consists of both the spatial factor (where) and the temporal factor (when) along with the post source which act as the attribute (what) of the object. Considering these factors the triad framework will be introduced into this data for further understanding. Using the triad framework we can form basic questions on the data such as: What made the user to post the picture from a particular place and at a particular time? Where the user is at the time of posting the picture and at what time is he posting it? Or questions like when the user is posting the picture with an ongoing disaster? With these kind of basic questions we can further understand the use of Flickr data when it is implemented in STC.

3.4. Case study on enviroCar data

enviroCar (2013) is a crowd sourced application for citizen science projects, in which volunteers are collecting the data to get insight into sustainable mobility. Citizens or simply the drivers collect the moving car data and share this data on the enviroCar platform, which can be later used for scientific analysis. Figure 3.2 shows how the data is collected on the enviroCar platform. The data available in this platform is collected with the help of a Bluetooth OBD-II adapter which is connected to the car engine. This Bluetooth adapter is interconnected to a GPS enabled Android smartphone running the enviroCar app. So the data collected through this app consists of both the space and the time attributes (tracks) because of the presence of GPS in the smart phone, along with the information like revolutions per minute and speed. The app estimates the car information such as fuel consumption, estimated fuel cost and CO₂ emission which can be used for environmental impact assessment.

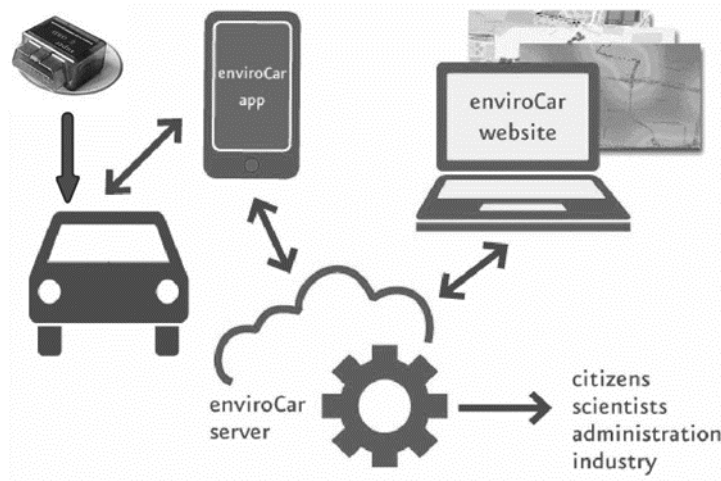


Figure 3.2: Representing enviroCar application

Since the enviroCar data consists of the time factor (when), the spatial factor (where) and the attribute data (what) triad framework can be applied for understanding the different spatio-temporal patterns of the data. The questions formed will be like: What is the fuel consumption level of the car in a known area, in a particular time? What are the places with most CO₂ emission under the known time and engine attributes? What is the speed of the car at a place in a particular time?

Since this is one of the case studies the next step is to look how this data can be visualised into STC. For getting insights into sustainable mobility we need to understand the spatio-temporal patterns in the data collected through this application. Hence STC is one of the options for this analysis because we can analyse both spatial and temporal factors at the same time.

3.5. Case study on OpenStreetMap data

OpenStreetMap (2004) is one of the crowdsourced data providers which allows the users to edit the present geo-information (GI) or to add the new GI. The GI is created and replaced by individual volunteers continuously. The main aim of the OSM is to create free and editable world map for all the users (“OpenStreetMap,” 2004). The volunteers contributing the data are expected to map the regions that they have visited or at least they must have some connections with the region of interest (Haklay, 2010). OSM maps in recent years are playing a crucial role in crisis management. Volunteers are publishing their results comparatively as efficient as the commercial and governmental data providers to help the people in crisis areas on time (Zook, Graham, Shelton, & Gorman, 2010). The areas with middle level of education, high population density, high rate of overnight stays, high income, high number of foreigners, and residents within the age group of 18 to 69 are more likely to involve in OSM (Arsanjani & Bakillah, 2014). But despite the continuous update cycle of OSM data it will never be complete.

In this case study, OSM data of Miyagi region from Japan and the Overijssel region from Netherlands are selected to know, how the data in a particular region is being contributed by the volunteers over time. The Miyagi region is the most affected region in Japan at the time of Tohoku earthquake which is followed by a Tsunami. This region is selected to understand the user participation at the time of the disaster when compared to the situation before and after the disaster. The Overijssel region from Netherlands is selected in order to understand the user participation in this region as it is not affected by any natural disaster like the Tohoku in Japan. As the OSM dataset has the timestamp as one of its attributes, it can be implemented in the STC for spatio-temporal analysis.

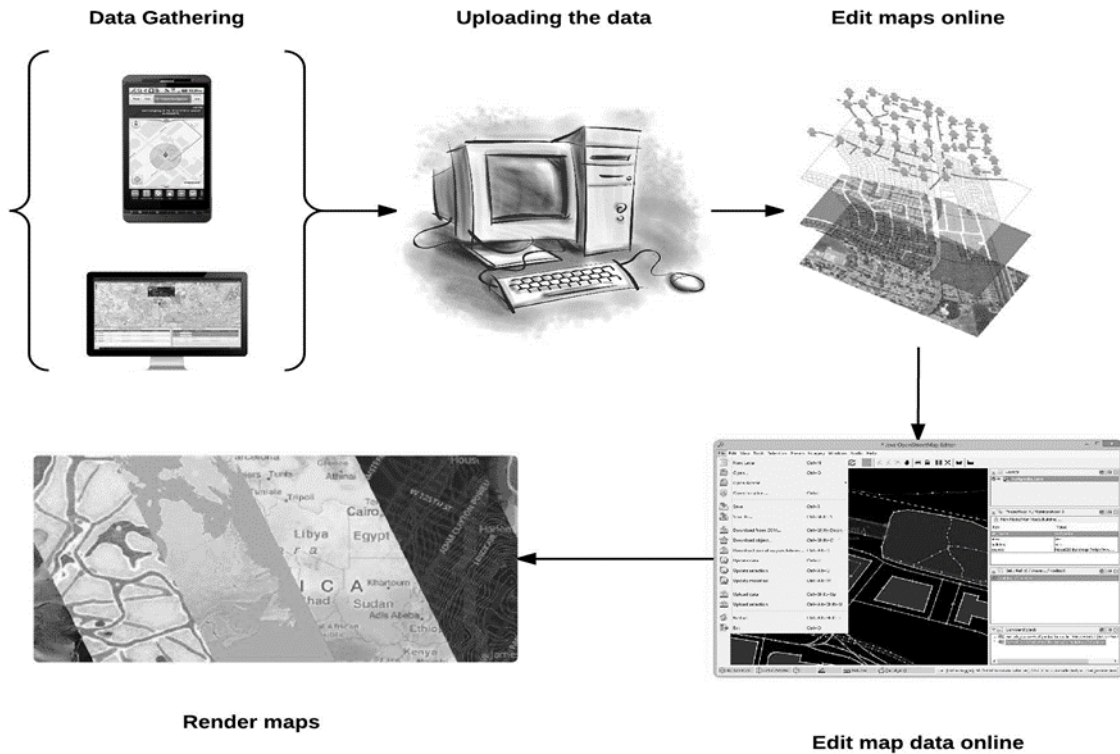


Figure 3.3: OSM data processing

Figure 3.3 explains the different steps to be followed on how the data is collected and made available for the users by the contributors. As the data consists of the timestamp which represents the time (when) of acquisition of the point along with the coordinate location of the point (where), this facilitates the spatio-

temporal pattern analysis of the data. It also consists of the user ID (what) of the contributor, from which we can extract the user details. As OSM data consists of three fields of data: where, what and when, it can be implemented in the triad framework. By applying the triad framework on the OSM data some basic questions can be formed such as: what made the contributors to contribute their data in that specific time, at what time period the development of the city is taken place mostly? In which areas the volunteers are mostly participating? What is the user participation at the time of disaster? With these kind of basic questions we can further understand the use of OSM data when it is implemented in STC.

3.6. Summary

This chapter introduced triad framework proposed by Peuquet and the structure of the three case studies: enviroCar, Flickr, and OpenStreetMap along with the importance of selecting these case studies. It is also explained that how these case studies can be interlinked with the triad framework and implemented in STC for further analysis.

4. VISUAL REPRESENTATION APPLIED TO CROWDSOURCED DATA

4.1. Introduction

In the previous chapters, the concepts related to this research such as, selected visualization methods for the visualization of crowdsourced geo-information and the crowdsourced data considered were described. In any visual representation, the functional environment consists of both interaction and visual representation functions. Hereby, it is important to know about the visual variables considered for visual representation. Hence, the next step of this research is to choose the appropriate visual variables for the visualization of crowdsourced geo information collected. For this reason, in this chapter a general understanding of the visual variables and their application's for different purposes will be discussed.

4.2. Analyzing the crowdsourced data collected

The data used in any visual representation consists of different components of information. It is important to know about the nature of the data, considering attribute information, time, and location to be able to select a proper visualization. Bertin (1983) primarily characterized the elements of information (data) into three levels of organization and named them as qualitative level or nominal level (when the categories of information are not ordered in universal manner), ordered (when the categories of information are ordered in a universally recognised manner), and quantitative is split into interval and ratio level (when the categories of information are used to specify the variation in countable units). See Table 4.1 for establishing the nature of the attribute information; when the attribute information is measured to define the attribute values of objects, it resembles that the characteristics of attributes are either quantitative or qualitative. Hence, the choice of representation used for visualization depends on the measurement scales (Kraak & Ormeling, 2011).

Measurement Scales	Explanation
Nominal Scale	All the attribute values are different in nature, without any superior classification (e.g. different languages)
Ordinal Scale	All the attribute values are different from each other, and there is a universally accepted way to classify them (e.g. temperature)
Interval Scale	All the attribute values are different, the distance between the individual measurements can be identified and can be ordered but cannot be compared (e.g. the currency difference between India and Europe)
Ratio Scale	The attributes in the ratio scale can be compared with each other, except that difference it is the same as the interval scale (e.g. number four can be said as twice of two)

Table 4.1: Measurement Scales (Kraak & Ormeling, 2011)

4.3. Visual Variables

Bertin is the first author to introduce visual variables for the representation of information on 2D maps. These variables played a prominent role in map understanding. According to Bertin (1983) there are two kinds of variables; namely, locational variables and retinal variables. The locational variables refer to the 2D data maps and graphs represented on (x, y) axes. Whereas, the retinal variables also known as the graphical or visual variables are the topics of symbolization. Bertin further divided the retinal variables into

six types; they are the size, shape, color, value, orientation, and texture (Figure 4.1). Later, researchers added new visual variables to these six visual variables, for example, the addition of crispness, resolution and transparency in the nine-variable model proposed by MacEachren (1994). But, these six visual variables presented by Bertin were considered to be the base of all the variables, so in this research we are going to discuss these six visual variables.

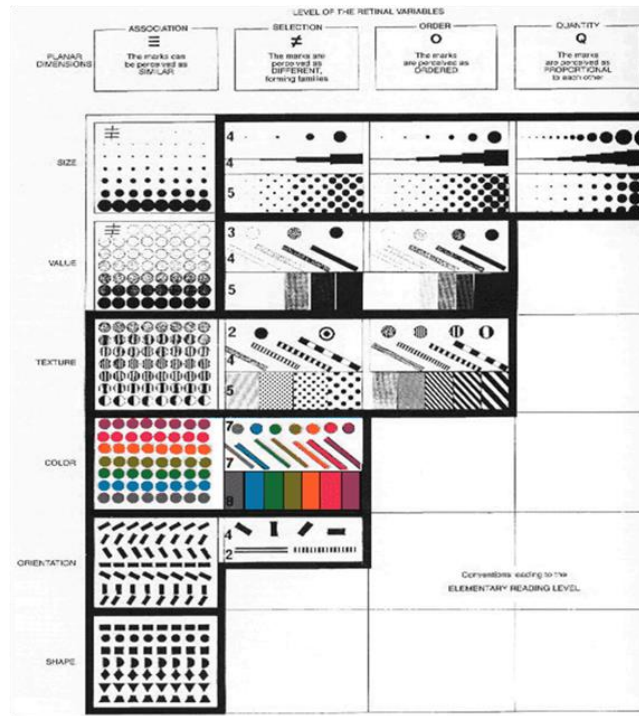


Figure 4.1: Retinal variables proposed by Bertin

While describing the visual variables of a map design, it is important to consider the dimensions of those visual variables. Monmonier (1993) and Bertin (1983) classified these dimensions into three types for easy understanding; they are the point symbols, line symbols, and areal symbols. Depending on the generalization of attribute data, these dimensions represent the zero, one, and two-dimensional features.

Bertin, further gave four perceptual approaches for questioning the characteristics of visual variables. For example, in visual representation size variation is very limited; if a size variable is combined with color variable, then color dominates the size. Hence, the perception of organization levels is much necessary. Each of these four aspects is explained in the table 4.2.

Organization levels	Explanation
Selective Perception	The variable used for the primary analysis and is used to differentiate a category from a group of categories is selective visual variable.
Associative Perception	If a variable is used to allow the grouping and equalization of all the categories of information during the perception of visualization, then the variable is called associative.
Quantitative Perception	If the difference between two categories among all the categories is interpreted numerically, then the visual variable is quantitative
Ordered Perception	If the visual variable representing the components of information are ordered in a universally accepted manner, then it is an ordered visual variable.

Table 4.2: Organization levels of Graphic variables

Selecting different visual variables for visualizing various aspects of attribute information can significantly influence the understanding of visualization method. For this reason, it is important to understand the appropriate use of the visual variables while creating any visual representation.

4.4. Application of visual variables to Space-Time Cube

The main objective of this research is to evaluate the usefulness of STC for visualizing the crowdsourced data; so, we consider the visual variables that can be visualized with-in the STC environment. The case studies in this research are considered in the context of point and line features, as described in the previous chapter. The options available for different visual variables in visualizing the point and line features is represented in figure 4.1. Also, the most common way to represent the spatio-temporal data in STC environment is through the use of Space-Time Path (STP) and the Stations. The STP represents the line features, and the stations represent the point features.

4.4.1. Application of visual variables to stations

Stations are the dimensionless point symbols that obviously need to have some shape, size and color for visualization. The centre of the station represents its location on a plane without occupying any space. As mentioned before, different visual variables can be applied to stations in STC environment. But, according to Bertin, point symbols have a limited number of classes to use, to keep it readable.

- **Size variation**

Size variation is used to represent the quantitative and ordered perception of the data. On an average, a human eye cannot distinguish the difference between two point's more than twenty steps, whose areas are in the ratio of 1 to 10 (Bertin, 1983). This limit in point data representation made Bertin (1983) to propose “natural series of graduated sizes” to visualize quantitative point data.

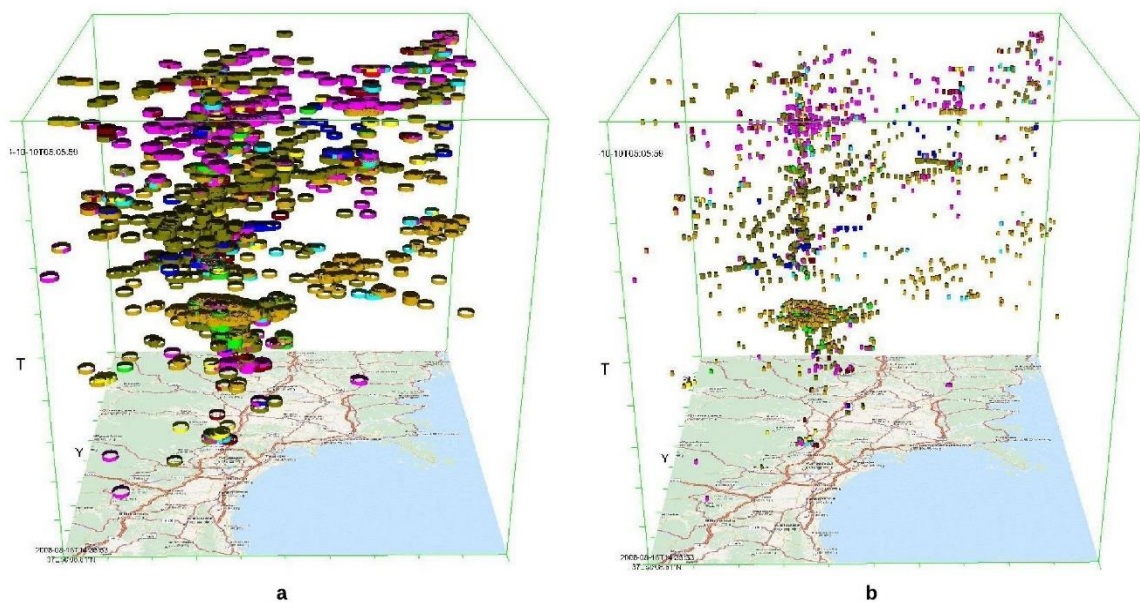


Figure 4.2: Size variable representation in STC environment

Figure 4.2 represents the advantage of the size variable for point symbols in an STC environment. Both images represent the same data, but image a, has more clutters than image b. Due to the change in the size of the attribute the cluttering of data is reduced. Once the cluttering of the data is reduced, there is an advantage in understanding the data more accurately.

- **Shape variation**

Shape is one of the useful variables in distinguishing the difference between the quantitative datasets. Bertin (1983) defined shape variation as an “element of “similarity” recognized in the shape which constitutes the stimulus for this variable”.

Figure 4.3 represents the usefulness of the shape variable for point symbols in STC environment. The different shapes (linear shaped points on the sides of the cube and circle shaped points at the center of the cube) of attributes in the cube represents the point data from different regions in Netherlands.

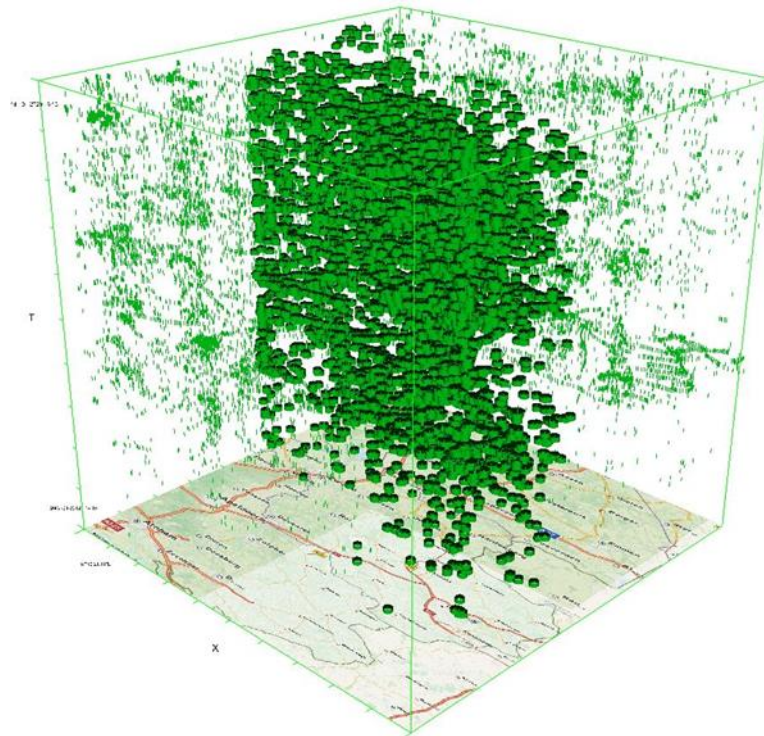


Figure 4.3: Shape variable representation in STC environment

- **Color variation**

Color is another useful variable for identifying the difference among the data sets. Bertin (1983) defined color variation as a “perceptual difference which can be perceived between uniform areas having the same value”.

Figure 4.4 represents the usefulness of color variable for point symbols in STC environment. The different colors of attributes in the cube represents the various categories of stations contributed by volunteers in OpenStreetMap. The use of color variable for differentiating the data helps in identifying different patterns within the data.

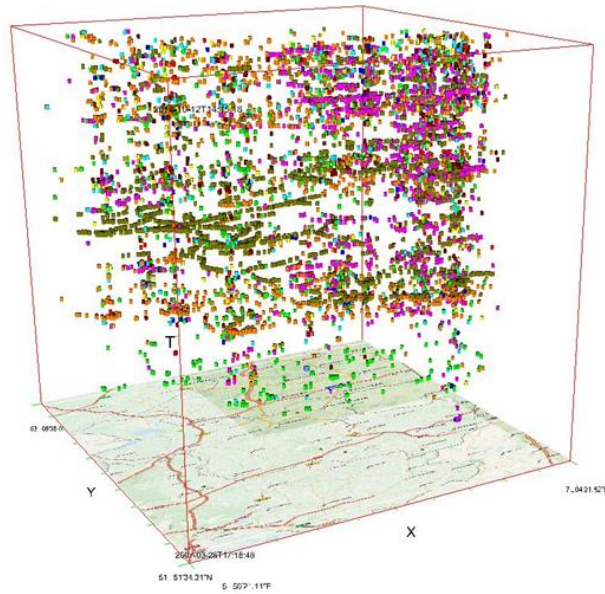


Figure 4.4: Color variable representation in STC environment

4.4.2. Application of visual variables to STP

STP's are also the dimensionless line symbols that obviously need to have some shape, size and color for visualization. The central axis of the path represents its exact location on a plane without occupying any space. In general the STP represents the moving track data, so the extent of the STP is fixed but the radius of the path can be varying. As mentioned before different visual variables can be applied to STP in STC environment.

- **Size variation**

As mentioned for stations, the size variable is one of the most commonly identified variation by human eye. In linear representation, the varying thickness of the line can bring the quantitative perception. As explained by Bertin, a natural series of graduated sizes can be used to differentiate the attributes with some limitations.

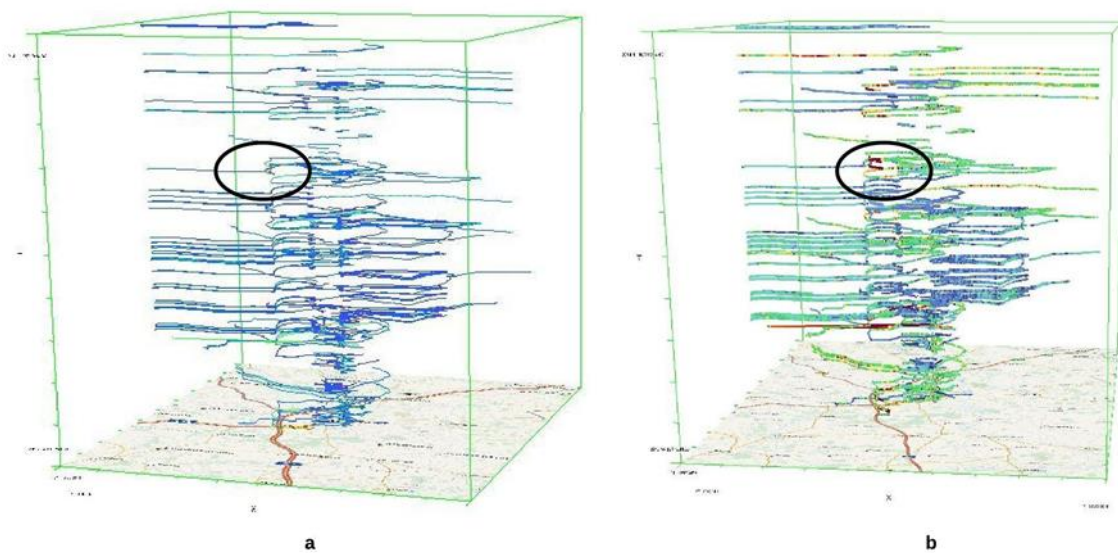


Figure 4.5: Size variation in STC environment

Figure 4.5 represents the importance of the size variable in STC environment. The two images “a” and “b” are implemented in an STC environment with the same data sets and same color variation to represent the CO₂ emission along the track but the size of the paths were different. In image a, the paths are much thicker than in image b. Due to the increase in the thickness of the path there is a possibility of identifying minute changes. At this moment we can see, there are some red spots in image a, which are not visible in image b.

- **Color variation**

As mentioned for station, color variation is a good option to represent the difference in path. Figure 4.7 represents different colors for different car (different car manufacturers) tracks. As it can be seen, it is the best possible way to describe the variation in car manufacturers.

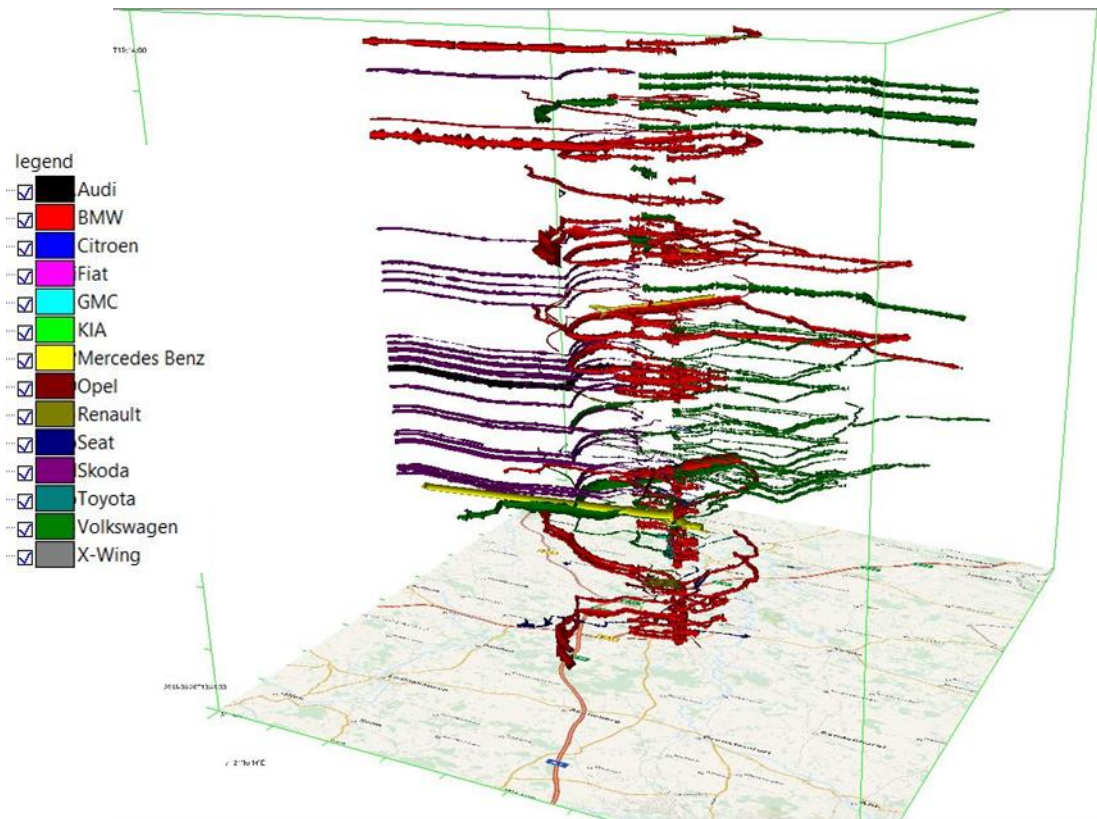


Figure 4.6: Shape and color variation in STC environment

- **Shape variation**

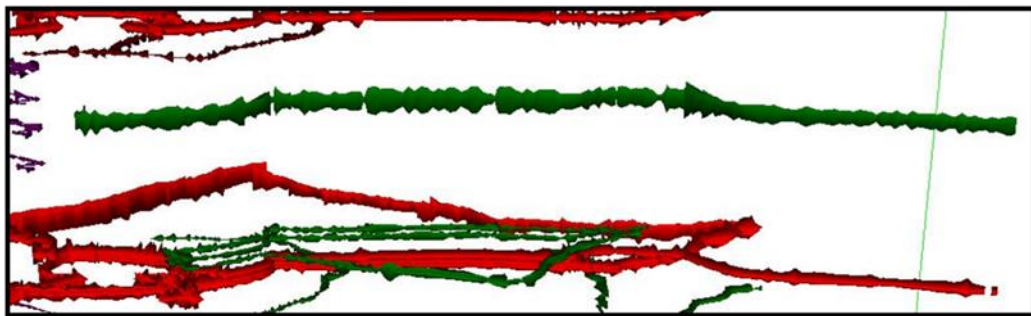


Figure 4.7: Shape variation of single track

As mentioned for the stations, shape is an important variable to be considered in visualization. Figure 4.6 represents the track data of different cars, and figure 4.7 represent the shape of a single track. The shape of the track represent the intensity of CO₂ emission along the track. Here, the triangle shape is used to differentiate the CO₂ values. The triangle shape has three edges,

which is the best possible way to represent the continuous change when compared to the other changes like the square or a circle.

4.5. Summary

This chapter is about the understanding the cartographic information, present visual variables and their application on Space-Time Cube for the better visualization of the crowdsourced data. These visual variables in STC are going to be implemented in the next chapter for spatio-temporal analysis of the case studies described in the previous chapter. In the next chapter, the conceptual framework and its implementation of crowdsourced applications will be discussed.

5. IMPLEMENTATION OF CROWDSOURCED DATA

5.1. Introduction

In Chapter-4, different options on how to visualize the data using different visual variables was introduced. Considering the technological developments and the options available, the prototype (Space-Time Cube tool) designed within the ILWIS environment is selected in order to explore the spatio-temporal patterns of crowdsourced geo-information. This chapter will discuss the implementation and the results of the case studies described in Chapter-3 using Space-Time Cube. This chapter will start with a brief introduction of the prototype and then follows with a discussion on the implementation of three case studies.

5.2. Visualization environment used

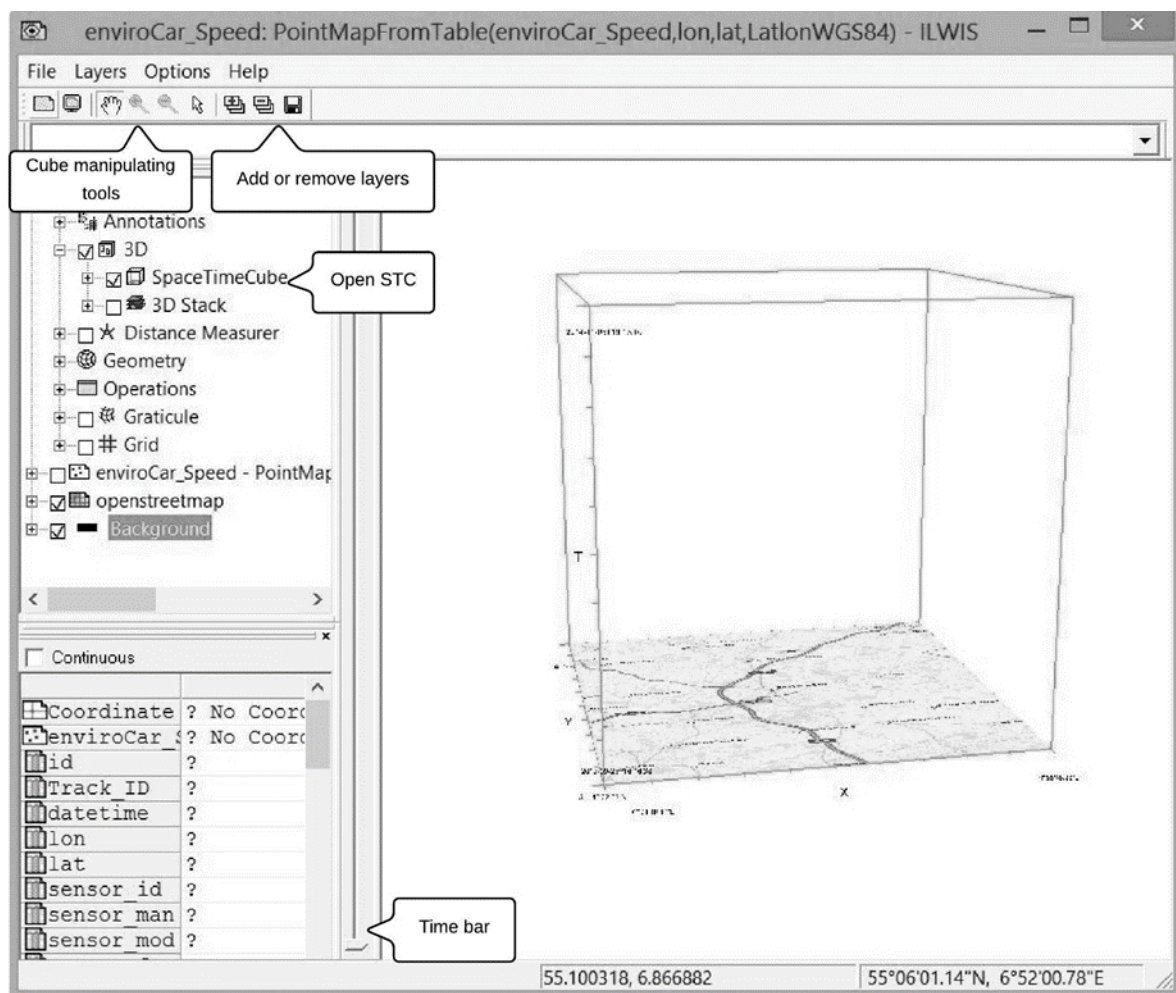


Figure 5.1: Space Time-Cube interface in ILWIS environment

ILWIS is a free desktop application, developed by ITC (Faculty of Geoinformation Science and Earth Observation), Netherlands for both vector and raster image processing. Space-Time Cube is a tool built in ILWIS application for spatio-temporal analysis under 3D visualization. Figure 5.1 represents the Space-Time Cube interface in ILWIS environment. The Space-Time Cube interface has different options like panning, zoom-in & zoom-out to manipulate the cube. There is an option to add different base maps

to the environment. The base map in the figure 5.1 is OpenStreetMap, which is mostly used as a base map in this research. There is a time bar in the interface, with which we can move the base map to different time periods.

5.3. Implementing the collected crowdsourced data

The objectives of the case studies used in this research and the collection of the data were described in Chapter-3. As mentioned, there will be three case studies used in this research for identifying the usefulness of Space-Time Cube in Spatio-temporal pattern analysis. Depending on the objective of the three case studies, data required for this research is collected. The information regarding the collected data and the implementation part will be explained in the following sections. Once the data is imported into STC, then spatio-temporal analysis will be performed by asking questions. These questions can be structured basing on the triad framework. However, majority of the questions that can be answered using the triad framework were not described in this research due to time constraints. Only some of the basic questions will be answered to show the usefulness of STC in spatio-temporal analysis.

5.3.1. Flickr case study implementation and analysis

As described in Chapter-3, this case study data is collected with an objective of studying the human behaviour at the time of disaster. The data required for this research is collected in the form of two layers (Flickr posts and Sandy Hurricane track). One layer of the data is collected through the Flickr API (URL 4). The data collected from Flickr API consists of owner ID, secret code of every post, title of the post, date taken, tags within the post, Latitude & Longitude values, unique place ID and a derived schedule of the day (morning, afternoon and evening). The data is collected from 21st October, 2012 (the starting day of Sandy hurricane and lasted until 31st October, 2012) to 9th November, 2012. There are totally 10,494 posts made by the users at that time period. Figure 5.2 represents the posts made by the users per each day.

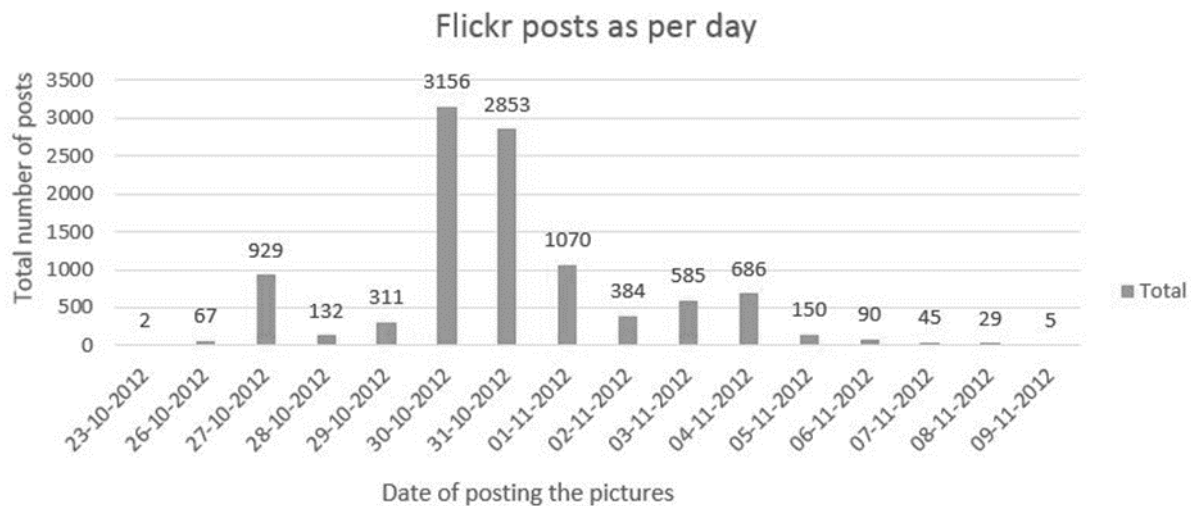


Figure 5.2: Graph representing the Flickr posts made by the users as per day

It can be identified from the above graph that, there are very few posts made by the users at the beginning, continued to increase in time and after some period of time the posts made by the users got decreased. A second layer of Hurricane Sandy track data is used in order to compare the posts of the users with the Hurricane movement. The Hurricane Sandy track data is collected from National Climatic Data Center (NOAA). This data act as the second layer for this case study. The Sandy hurricane track data is a continuous data consisting of Latitude & Longitude values, Wind pressure, wind speed, and the date taken

values for every point. The Sandy Hurricane was started on 21st October, 2012 and ended on 31st October, 2012.

For loading the collected data into the Space-Time Cube first we need to import the data into the ILWIS format. Once the data is imported, the Flickr posts posted by the users in the United States region is displayed in 2D format. Figure 5.3 represents 2D view of the Flickr data in ILWIS environment with OpenStreetMap background.

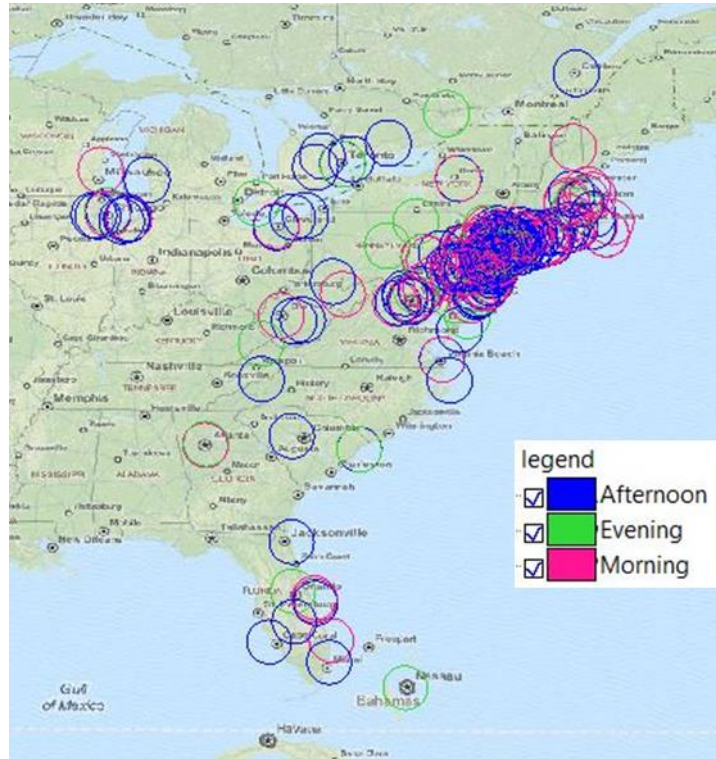


Figure 5.3: 2D view of Flickr posts

Once the data is loaded into ILWIS environment, the Space-Time Cube can be activated by checking the 3D option and the Space-Time Cube option within it. In the Space-Time Cube the data can be represented in different ways, by changing the display methods. The continuous data can be represented with the help of Space-Time Paths and the point data can be represented with the help of Stations. Figure 5.4 represents the Dialog interface to load data into the Space-Time Cube, in which the continuous data like Sandy Hurricane data is loaded with the help of Space-Time Path (STP) and the individual point data of Flickr posts were loaded with the help of Stations. The size of the Sandy track data is given according to the pressure of the wind.

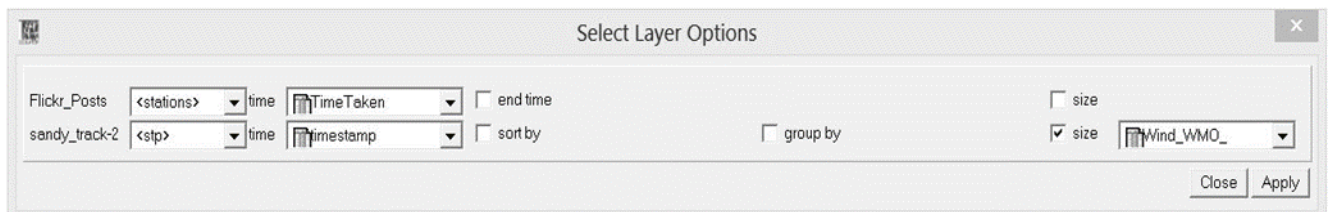


Figure 5.4: Interface to load data into Space-Time Cube

Once the data is implemented into the Space-Time Cube the variables used to represent the data are changed, as described in Chapter-4. The size of the stations is reduced to overcome the problems with

cluttering of data, different colors are given to the stations to differentiate the posts made by the users from morning, afternoon and night. The differentiation of posts will help in identifying the most frequent hours of user participation. The different posts made by the users are ordered according to time by placing the first post at the bottom of the cube to the last post at the top of the cube on their geographic location. Coming to Sandy Hurricane path, the size and color of the path represents the wind pressure. The STP is designed in such a way that it connects individual continuous points in different locations in the order of time. The same rule is applied to represent the Sandy Hurricane path as well. Figure 5.5 represents the Flickr posts made by the users along with the Sandy Hurricane path. The legend used to represent the data is also given in the figure.

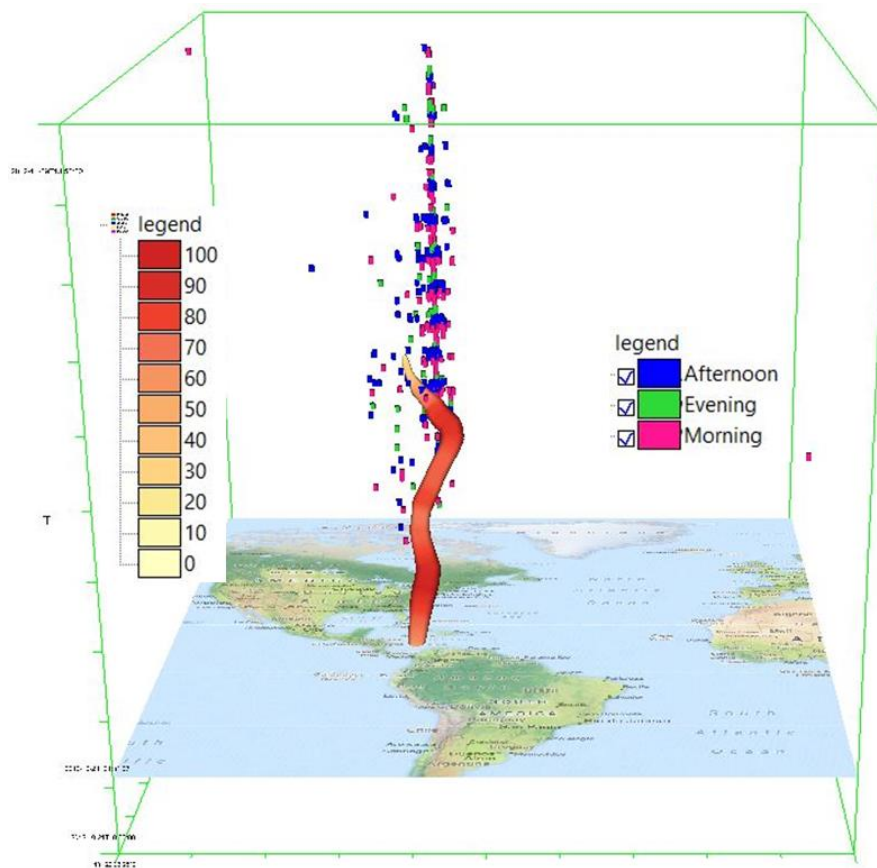


Figure 5.5: Implementation of Flickr case study in Space-Time Cube

As mentioned before, only the basic questions formed by the using the triad framework will be explained due to time constraints. For example, using STC we can answer questions like what made the user to post the pictures when the time and location of the post is known. Figure 5.6 represents the spatio-temporal analysis conducted on STC for answering this question. Looking at Figure 5.7, we can identify the increase of posts made by the users when the hurricane moves closer to the city. When the hurricane is formed at the middle of the ocean there are no posts made by the users, once the hurricane started to move towards the New York City coast the user started to post their posts, and once the Hurricane entered the city the posts made by the users increased, that too more posts near to the hurricane. This proves that the users are trying to spread the information on effect of Hurricane in the city through their posts. It is also interesting to know from Figure 5.5, how the people are posting even after the impact of the Hurricane is reduced. Another important fact is most of the posts made by the user are at morning and afternoon times. Maybe, morning and afternoon are related to travel to and from work hours in which the users are immediately posting what they see. This will help the other users of Flickr to know the information regarding

other places and take precautions before going to that place. Similarly, the users can apply same methods in identifying new spatio-temporal in STC.

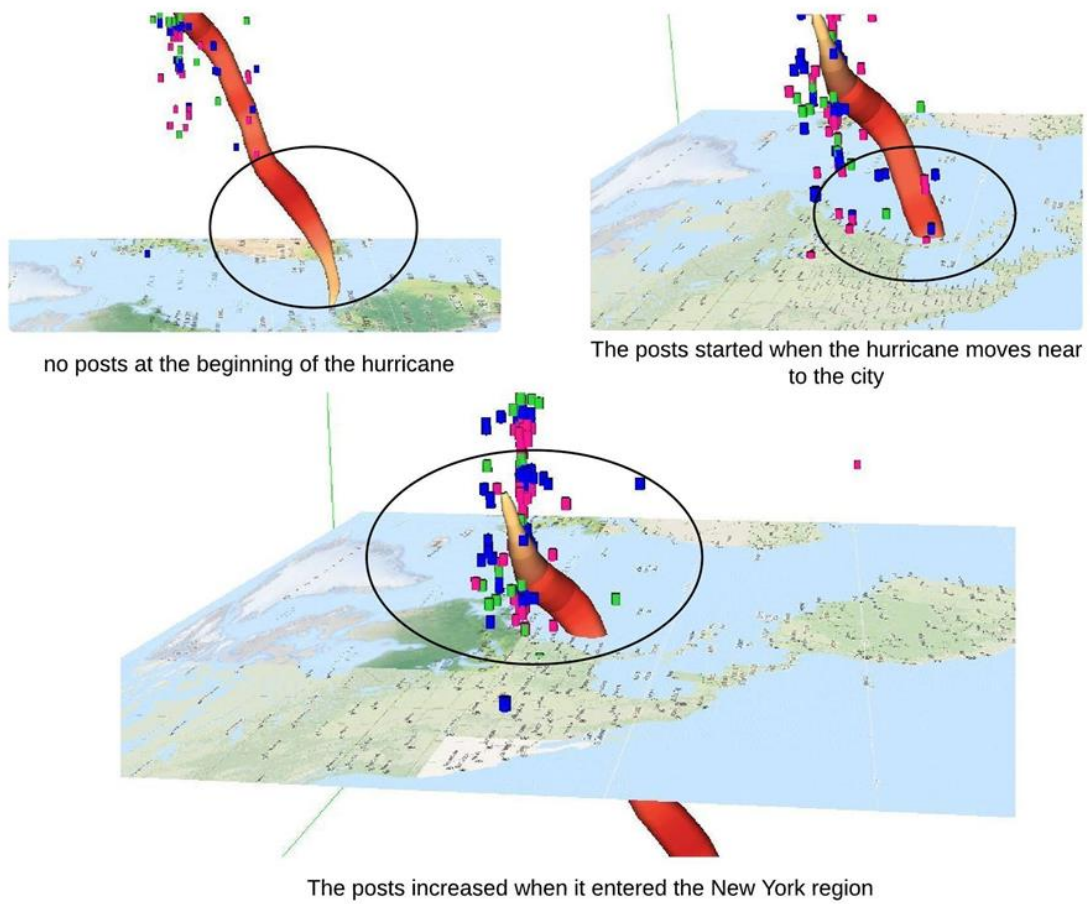


Figure 5.6: Spatio temporal analysis on Flickr data

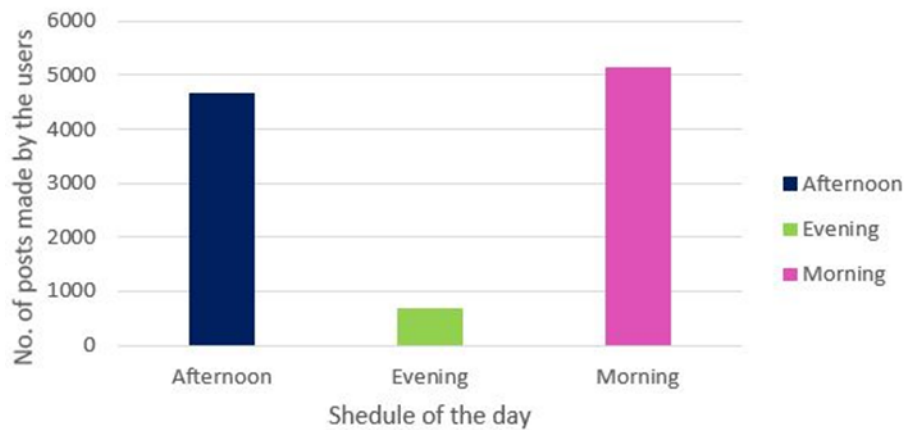


Figure 5.7: Graph representing the no. of posts made by the users

5.3.2. enviroCar case study implementation and analysis

As described in Chapter-3, enviroCar data is collected with an objective of conducting spatio-temporal analysis on the emission of CO₂ within STC. The data required for this research is collected from the enviroCar API and consists of different attributes for every point like the speed of the car, CO₂ emission values, collected date and time, Latitude & Longitude values, Fuel type, car manufacturer, Track ID and the sensor ID. The data collected is in between the time period of 25th September, 2013 to 5th December, 2014. This is an enormous dataset (nearly consisting of 540 different tracks) consisting of different attributes. There are a lot of tracks with missing attributes values, such as missing CO₂ emission values. Such data is not useful for analysis, so it is removed in data mining and only the tracks with all attributes are considered for analysis. After data mining, the data is left with ten different types of car manufacturers. Figure 5.8 represents the different car manufacturers and the point data available for each car manufacturer.

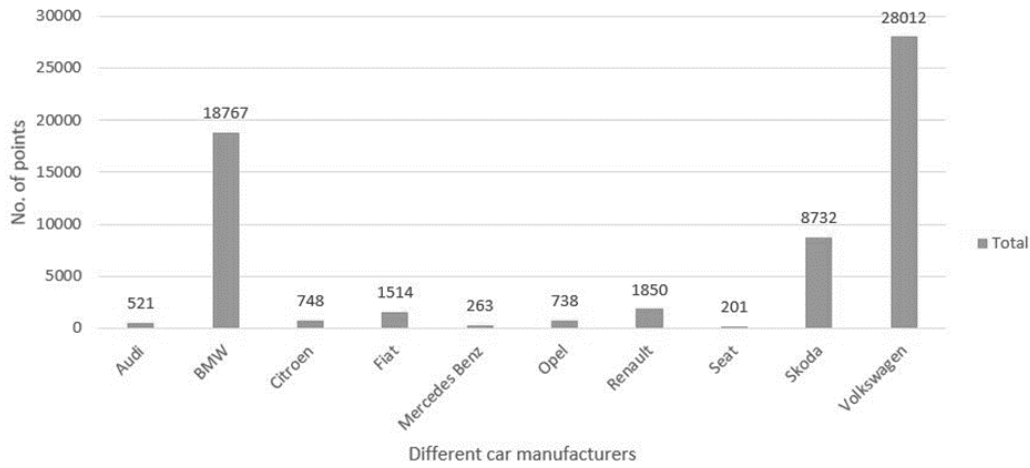


Figure 5.8: enviroCar data classification

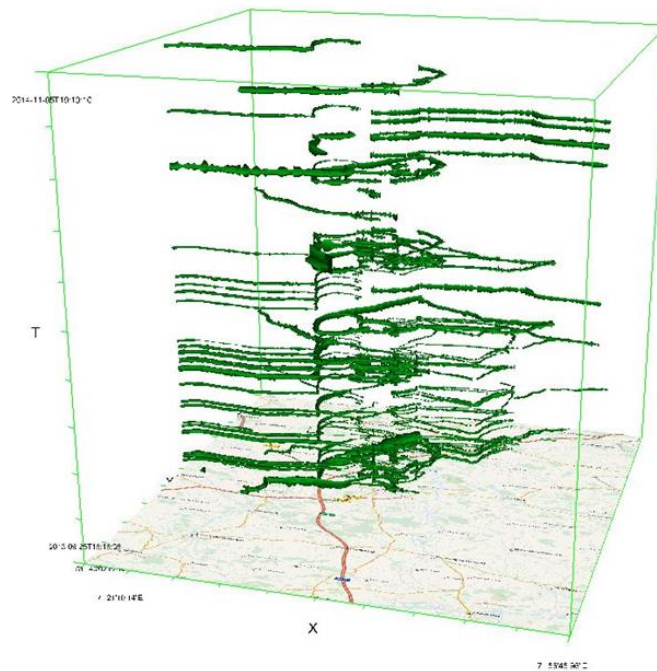


Figure 5.9: enviroCar data in STC environment

Once the data is prepared, it is imported into ILWIS software and loaded into the Space-Time Cube. As the data is continuous moving data, it is represented by using the concept of Space-Time Path (STP). The size of the path represent the CO₂ emission of the car. Figure 5.9 represents the primary implementation of enviroCar data into STC. OpenStreetMap is used as the base map to represent the spatial coverage of the data. Now, different visual variables can be applied to this visualization for the spatio-temporal analysis. For example, by separating the car tracks in figure 5.10 into two different colors (gasoline and diesel), we can compare the difference in the emissions of CO₂ for both fuel type cars. In Figure 5.10 the tracks are classified depending on the fuel type. The red color tracks represent the diesel cars and the yellow color represents the gasoline cars. From representation it can be identified that the gasoline cars are emitting less CO₂ values than the diesel cars. It must be noted that the emission of CO₂ depends on lot of varying factors (car type, engine attributes, fuel consumption & etc.), but for this representation the interpretation is made on the basis of visualization.

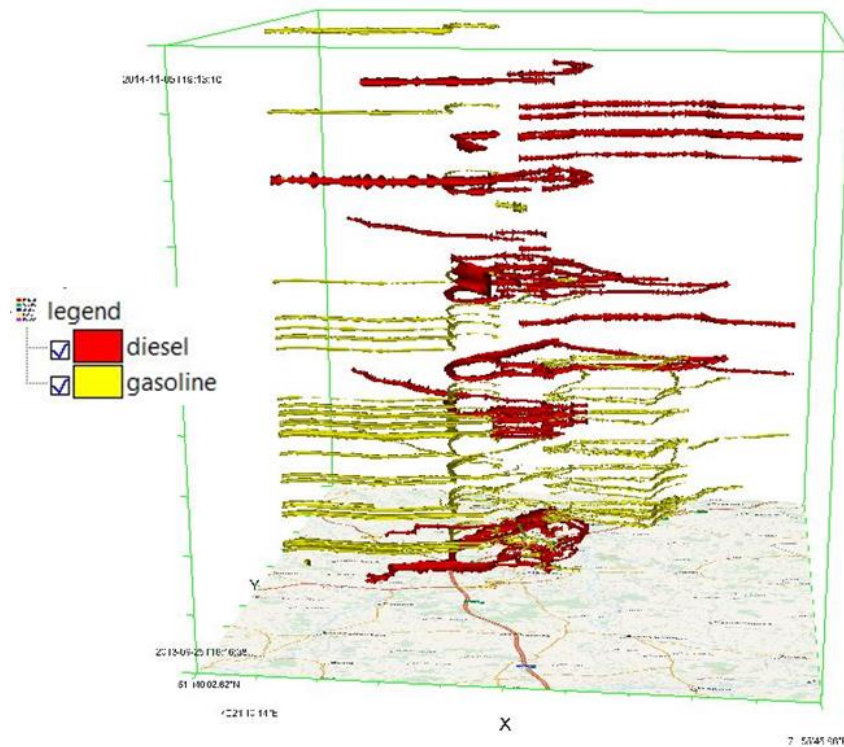


Figure 5.10: enviroCar tracks classified on the basis of fuel consumption

The Triad framework can be used to formulate some basic questions on enviroCar data to identify the spatio-temporal patterns. The fundamental question like “where + when → what” of the Triad framework helps in identifying different patterns in the data. For example, we can formulate a question on; how does the CO₂ emission (what) vary at a four way junctions (where) in different time periods (when)? Answering this question might lead to the identification of new spatio-temporal patterns.

For answering the above question spatio temporal analysis is conducted in STC using different visual variables. Figure 5.11 represents the analysis conducted on STC. Colour of the track represents different car manufacturers and the size represents the emission of CO₂. In order to answer the above question a four way junction is identified and the analysis is conducted. First the analysis is conducted on one single track. It can be seen from the image that the emission of CO₂ decreases whenever the car takes the turn at the junction. The point to be noted here is the emission of CO₂ is directly related to the speed of the car. If the speed increases the emission of CO₂ increases. Due to this reason, whenever the car reaches the junction, speed of the car decreases which reduces the emission of CO₂ and once it crosses the junction the speed increases and CO₂ emission also increases. When this phenomenon is compared to rest of the tracks passing through that junction gave the same result as the first track. This phenomenon can be applied to different application like the identification of speedbumps and traffic signals. Due to time constraints, only one example is explained in this section.

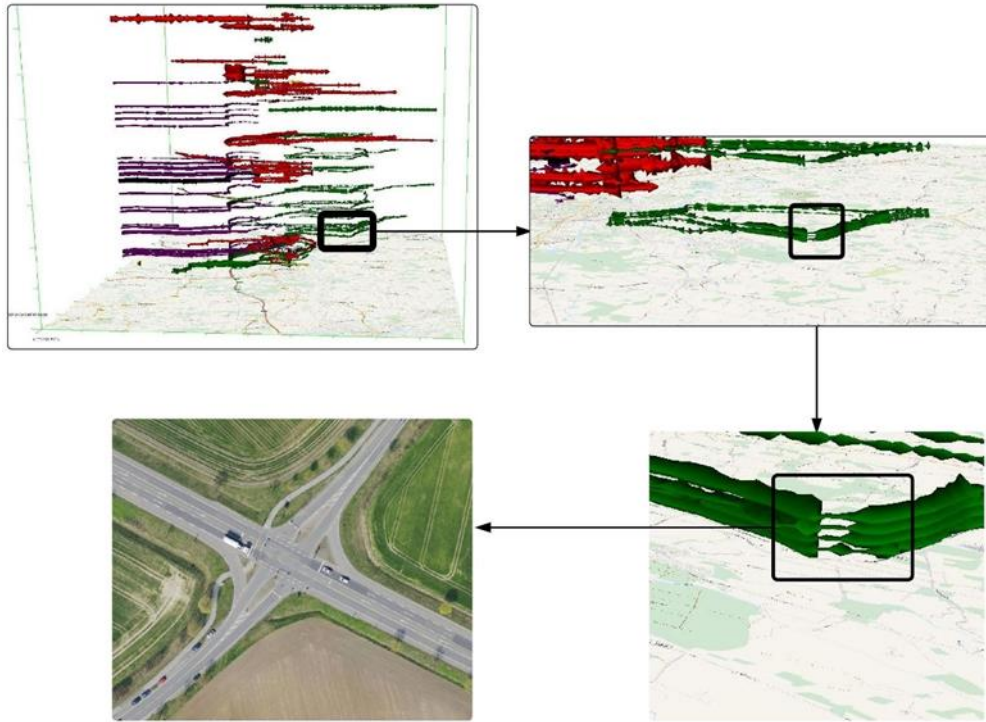


Figure 5.11: Spatio-temporal analysis on enviroCar data

Space-Time Cube is also useful in identifying the anomalies in the data. By using the visual variables we can display the data for analysis, and with the help of cube manipulating tools we can move the cube and identify the anomalies which are different from other attributes. Figure 5.12 represents one of the anomalies (high CO₂ emission) identified in one side of the cube. One of the main advantage of the STC is once the any anomalies in a track are identified we can compare them with the other tracks and come to a conclusion at a time. There are many other ways to identify the spatio-temporal patterns in enviroCar data within STC, but considering the time only some examples were discussed in this section.

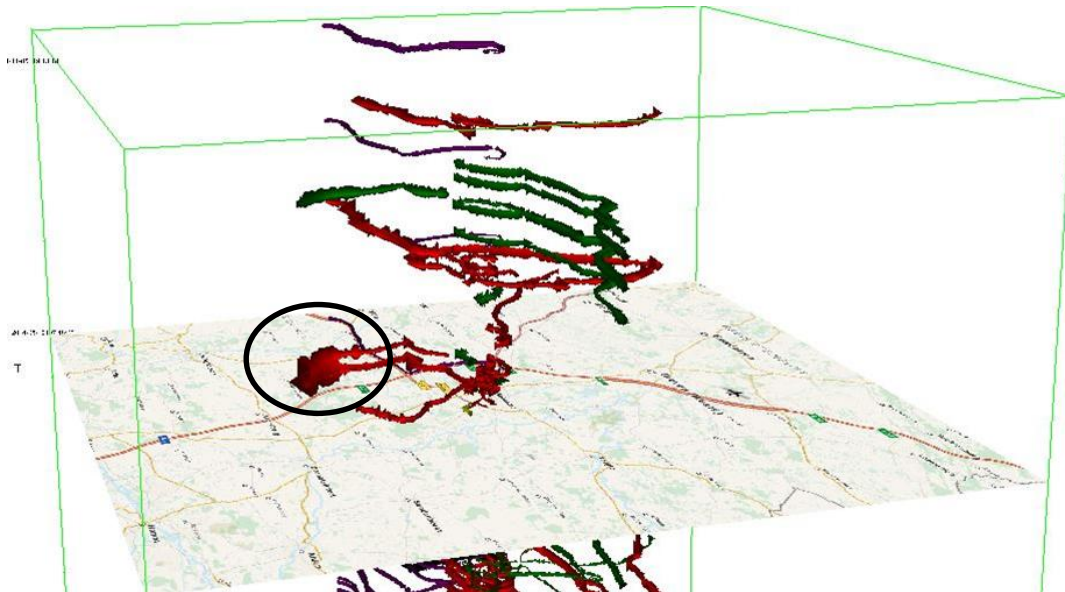


Figure 5.12: Identification of anomalies in STC

5.3.3. OpenStreetMap case study implementation and analysis

As described in Chapter-3, OpenStreetMap (OSM) data is collected with an objective of understanding the nature of the OSM contributors. Only the point data can be implemented into STC, so all the data (Miyagi, Overijssel, Haiti, Tanzania regions data) required for spatio-temporal analysis of this case study is extracted from the OSM dump database. The data consists of point Latitude & Longitude values, timestamp of acquisition, point type and region information. There are almost 255 different point types contributed by the OSM users, which is hard to visualize in any visualization. So for the easiness of visualization, the different point types were classified again into nine fields depending on the type of information. They are classified by combining all the Bars, restaurants and fast-food centres into one field (Bars_Restaurants_FastFoods), all the parking and rental centres into one field (Parking_Rental), all banks and ATMs' into one field (Banks_ATMs), all the governmental service points and the commercial service points into one field (Governmental_Commercial_Services), all transportation related points into one field, all the shops and supermarkets into one field (Shops_SuperMarkets), all the recreational places into one field, all the educational institutes into one field and the rest into one field (Others). Figure 5.13 & Figure 5.14 represents the point data contributed by different users in different years. Only the years with high user participation are considered for representation. Different colors represent different months in a year.

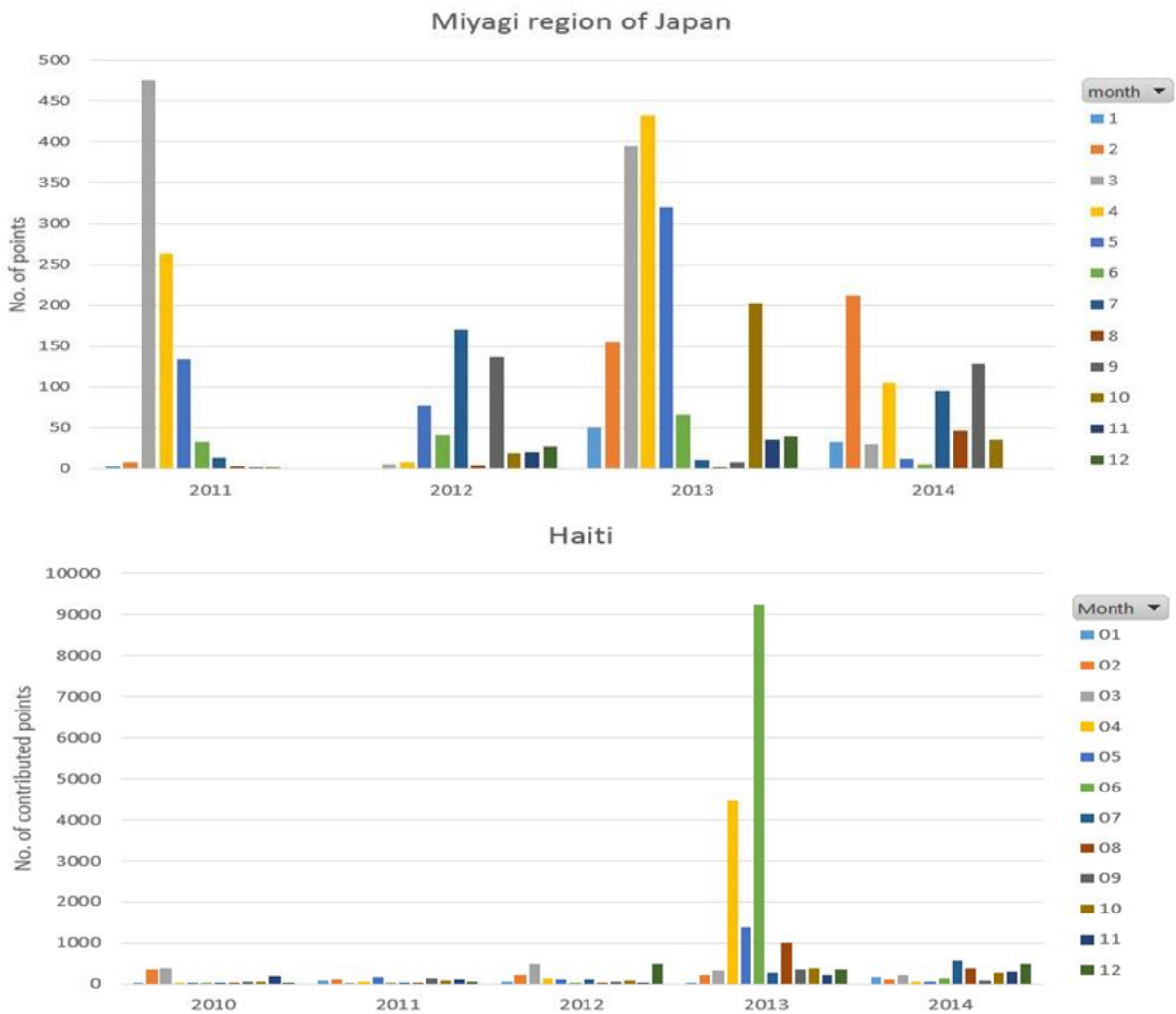


Figure 5.13: OSM data collected from different years in disaster effected regions

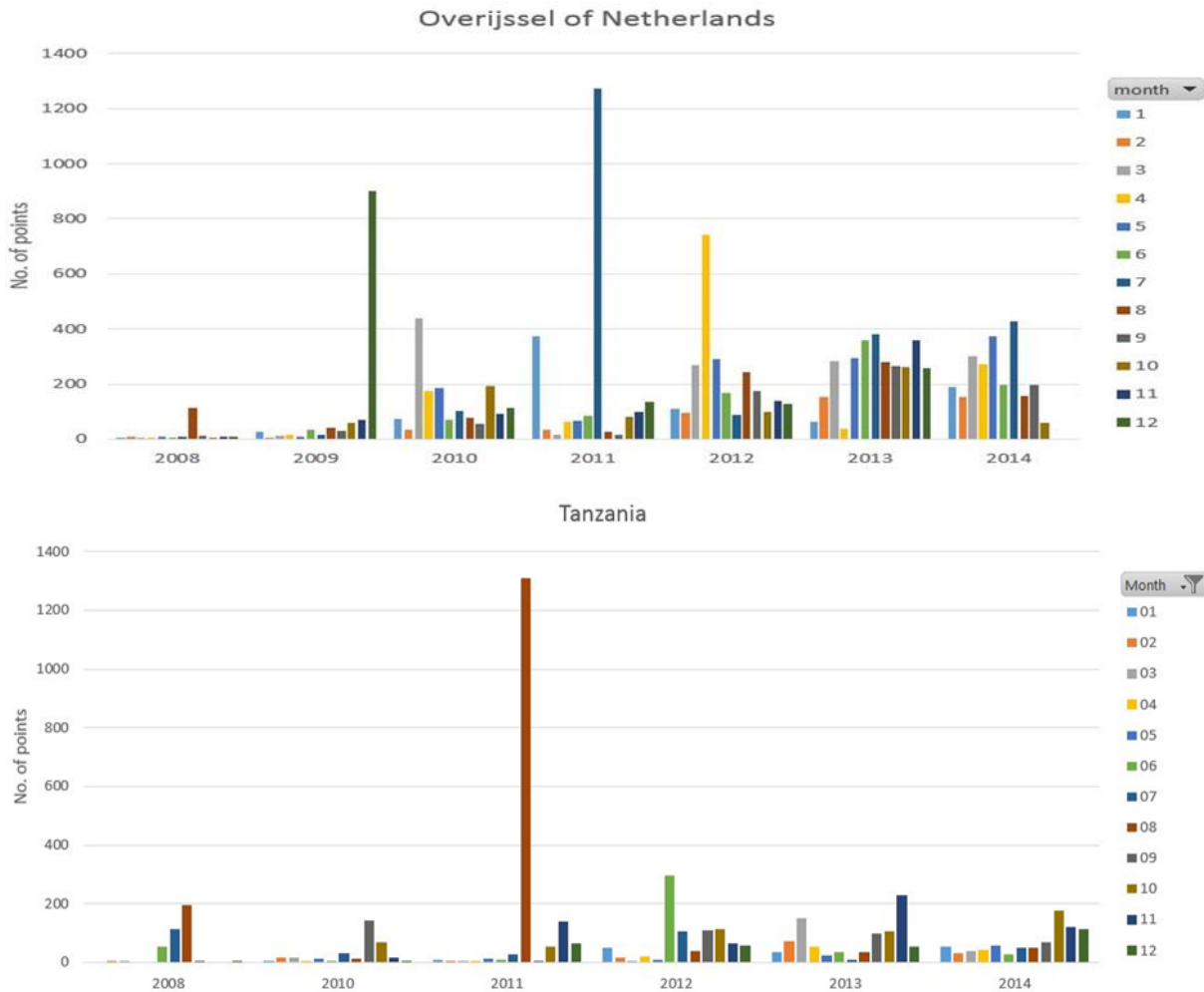


Figure 5.14: OSM data collected from different years in different regions

Separate shapefiles' are prepared for different regions and are imported into ILWIS software. Being individual point features, all the points were loaded into STC by using the concept of stations. OpenStreetMap is used as the base map to represent the spatial coverage of the data. Different visual variables can be applied to this visualization for the spatio-temporal analysis. The colour variable is used to differentiate the 9 different types of point data, and the size variable is used to reduce the cluttering of the data. For the easy understanding of the analysis same legend (Figure 5.15) is used for analysis on all the regions considered in this case study. In the coming sections every individual region will be described then the results will be compared with the rest of the regions.



Figure 5.15: Legend of the OSM data

5.3.3.1. Miyagi region

Miyagi region of Japan is the most effected region at the time of Tohoku earthquake in March 11th, 2011. The reason for selecting this region as one of the case study is to analyse the OSM mappers' contributions at the time of disaster. Figure 5.16 represents the point data of Miyagi region in STC environment. Different colors represent the different point types contributed by the OSM mappers. The users started to contribute their data since 2008, but their participation is very limited until 2010. However, at the time of Tohoku earthquake the user participation increased enormously. They first contributed the information related to transportation facilities, then the information related to parking areas, and later on they started to share the information related to recreational places. The clutter identified in Figure 5.16 represents the contributed data at the time of disaster. The users started to contribute the data within 24 hours after the occurrence of earthquake, this indicates the active participation of the users at the time of disaster. From the figure 5.16 it can also be pointed out that the users are providing the information related to transportation and tourism. After the disaster the participation of the users had almost stopped for 7 months, this might be because of the constructions that took place after the disaster. Also after the disaster, there are spatio-temporal patterns identified in the data, mainly the patterns of same type of data (transportation, educational institutes, etc.) contributed by the users at a time. These form a horizontal planes that can be identified by manipulating the cube. This shows that the mappers are interested in mapping the fields with specific notation.

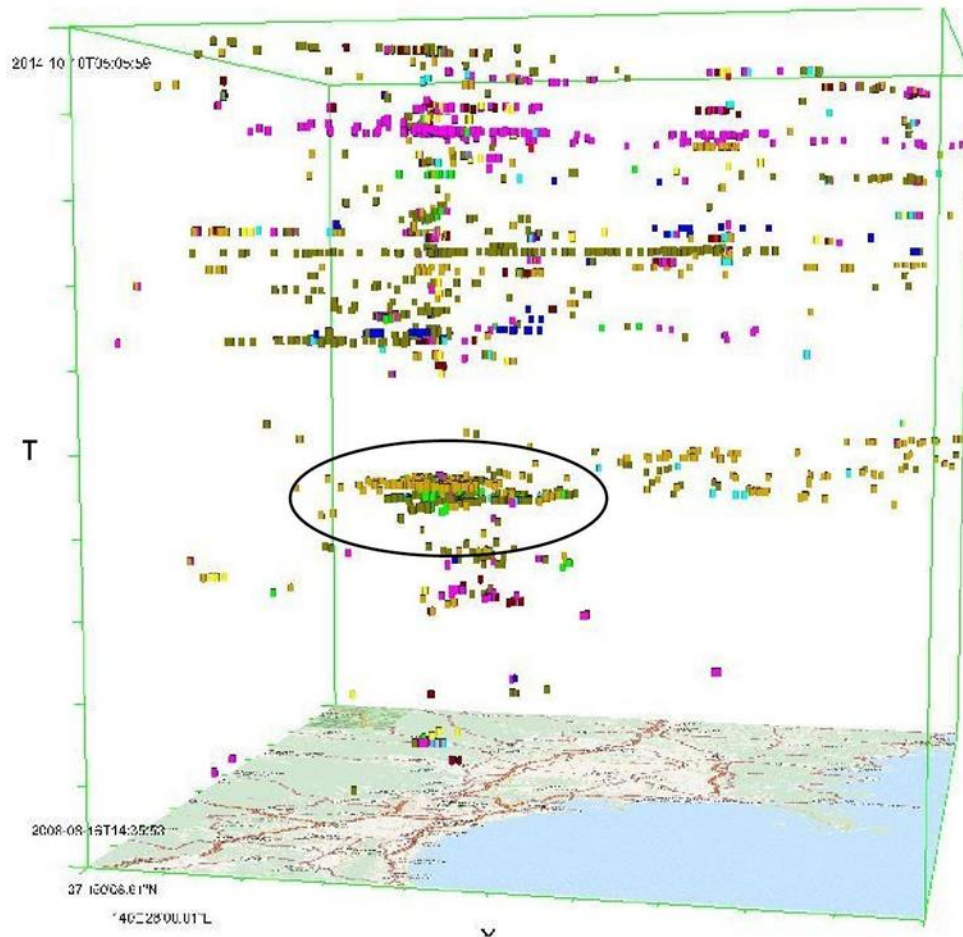


Figure 5.16: OSM contributed data of Miyagi Region in STC

5.3.3.2. Haiti region

Haiti region is also effected with an earthquake in January 24th, 2010. This region is also selected to analyse the OSM mappers' contribution at the time of disaster. Figure 5.17 represents the point data of Haiti region in STC environment. The same legend is used to represent the data in this visualization as well. Unlike Miyagi region, there is no active participation of the OSM mappers in Haiti region before the disaster. But, after the disaster, the mappers started to contribute their data using OSM. At the beginning, the mappers started to spread the information regarding the hospitals and the government help line centres. Again in the same way, OSM mappers stopped to contribute the data after the disaster for another 6-7 months. Most of the users contributed their data in the Port-au-Prince city, which is the capital and the largest city of Haiti. From figure-17 we can also identify some horizontal patterns of same type, which shows the mappers interest in contributing the data with specific notation. In the recent years there is a study increase in the users' contribution, which also shows the growth in crowdsourced application usage.

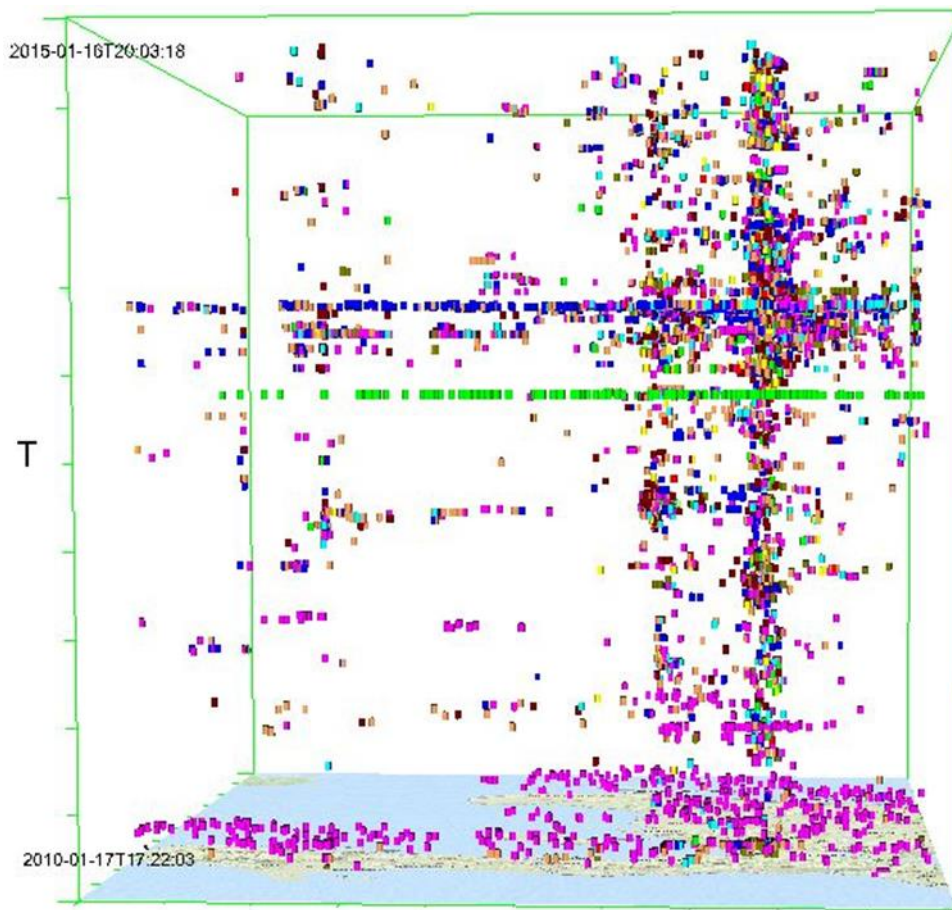


Figure 5.17: OSM contributed data of Haiti region in STC

5.3.3.3. Overijssel region

Netherlands is one of the developed countries. Overijssel region of Netherlands is selected in this case study as it will be hard to analyse the complete country at a time. Figure 5.18 represents the point data of Overijssel region in STC environment. Being a part of developed country, the user participation is high in this region when compared to the other regions of this case study. The users of this region started to contribute their data since 2007. At the beginning the users' contribution is very less, which is the same with the rest of the regions. However, the mappers' contribution started to increase with time. The development first took place at the city centres and started to spread along the entire city. From horizontal patterns, we can identify that the data is contributed by one specific user. It can also be note that there are a lot of

horizontal patterns recognized in this region, which shows that, most of the contributors in this region are collecting the data with a particular theme to contribute. Unlike the Miyagi or Haiti regions there is no specific gaps between the data contribution.

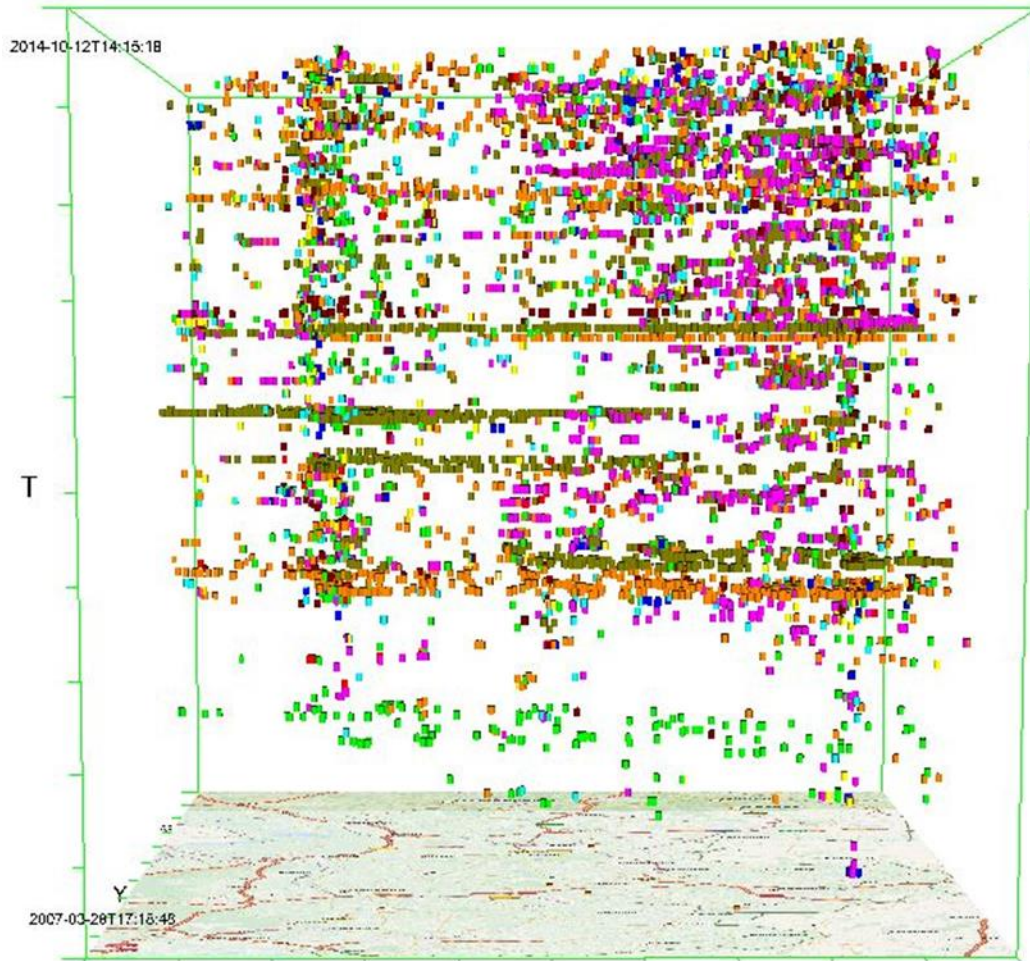


Figure 5.18: OSM contributed data of Overijssel in STC

5.3.3.4. Tanzania region

Tanzania of Africa is one of the developing country of this century. This country is selected as one of the case study region of this research to identify the mappers' interest in this region. Figure 5.19 represents the OSM mappers' contribution in Tanzania. Being a developing country the OSM mappers contribution is less when compared to the rest of the regions in this case study. The most contributions are found at tourist places in this county. Unlike other regions the horizontal patterns are also very less in this country data, which shows that most of the users mapping in this country are not contributing the data with any specific theme. But the user participation is increasing from time to time, which signifies that the user participation will increase in future.

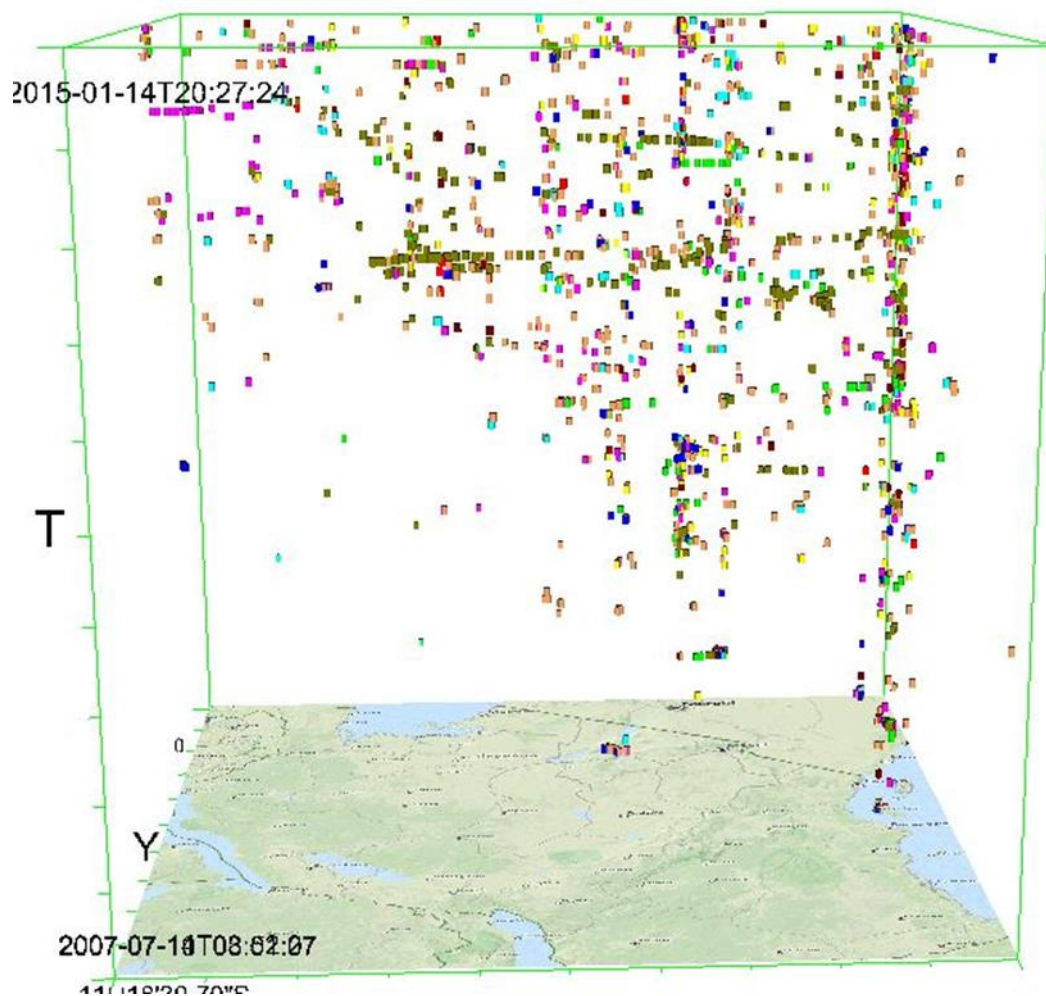


Figure 5.19: OSM user contributed data of Tanzania in STC

There is an increase in the user participation in all the regions considered in this case study, which signifies that the increase in the OSM user community. Also from the Japan and Haiti analysis the usefulness of the crowdsourced applications such as OSM at the time of disaster can be noted.

5.4. Summary

This chapter has explained how the data considered in the three case studies is implemented into STC and followed by explaining the preliminary analysis conducted on the implemented data. The use of visual variables that are described in the previous chapter are illustrated with examples in this chapter. Usefulness of the Triad framework in formulating the basic questions to identify the spatio-temporal pattern is described with the help of one example for each case study. The visualizations are designed and implemented in such a way that they will demonstrate the usefulness of Space-Time Cube in the spatio-temporal pattern exploration. The visualizations described in this chapter, mainly the visualizations on the enviroCar case study will be evaluated to verify the suitability levels in the next chapter.

6. EVALUATION

6.1. Introduction

In the previous chapters the design of the visual environment and its implementation were discussed. To test whether the solution offered in this research work in the context of spatio-temporal pattern analysis, a usability evaluation is necessary. Systematic usability analysis also helps in identifying the strengths and weaknesses of the Space-Time Cube (STC). Considering the availability of many new cartographic visualization techniques, the issues of usability and fitness for use are getting increasing attention. Kveladze et al., (2013) covered these issues in their research by conducting usability analyses for the STC. Usability is defined as “The extent to which a product can be used by specified users to achieve specified goals with the effectiveness, efficiency, and satisfaction in a specified context of use” (URL 5). For this reason efficiency, effectiveness, and satisfaction are considered to be the most used measures in finding the usability of a visual representation.

This chapter defines the usability measures considered in this research, with a clear description of the usability analysis conducted for the targetted group, the results of the usability analysis and followed by a discussion and a summary of the chapter.

6.2. Selecting an evaluation method

According to ISO 9241-11 standards the effectiveness, efficiency and satisfaction are the three primary aspects of usability. They are defined as:

Effectiveness: “The accuracy and completeness with which users achieve specified goals.”

Efficiency: “The resources expended in relation to the accuracy and completeness with which users achieve goals.”

Satisfaction: “The comfort and acceptability of use.”

These three measures of usability can be applied to the evaluation of any visual representation. In this research, the usefulness of Space-Time Cube for representing spatio-temporal patterns in the context of crowdsourced application (enviroCar) have been tested against these three ISO usability measures. For obtaining the results of these three measures of usability, there are different evaluation methods to follow.

The evaluation of the visualization results were necessary to develop the existing ones and gather new ideas. In addition, it helps the designer to create a better robust visualization. Focus groups are one of the methods of evaluation and are mainly targeted in the marketing field, in order to get insights into customer preferences and ideas. The word “focus group” concentrates on a group of users with a particular size, composition, procedures and are entitled to be specialized in a particular field. Video conferences, isolated web evaluation, and groups gathering are the most common form of focus group analysis. Focus group collects observational and verbal data. On the other hand, along with the focus group evaluation, there are other usability evaluation methods like think aloud, group discussion, questionnaires and eye tracking evaluation methods.

Questionnaires are one of the evaluation methods in usability analysis. In this method, the users are addressed with some questions and their responses will be recorded. Conducting interviews is also similar to questionnaires method, but it has some advantages over it. The interviews are much lighter than questionnaires and have high response rates (J. Nielsen, 1993) as the interviewer can rephrase the questions and explain them in depth. But, some of the answers we get from the participants might not be scientifically accepted, and there is a probability of getting arbitrary responses to the questions they cannot answer.

Think aloud is another type of evaluation methods in which participants are expected to think out loud in front of a system while executing a task (J. Nielsen, 1993). Sometimes, the actions of the participant will be recorded to make it more sophisticated. The problem with think aloud method is, it needs a lot of coding to understand the meanings of actions and verbalization of participants. These issues make think

aloud method to be more extensive and expensive method as it requires videotapes for recording the actions and a high-end computer to study and interpret the results (J. Nielsen, 1993).

Eye tracking method is another type of evaluation method in which the participants are monitored with an eye tracking device and a system. This is an expensive method in which the eye tracking device will record the eye movement of the participant on the computer screen. This method also needs a lot of coding to integrate and analyze the results. The combination of the think aloud method and the eye tracking method will help in getting better results.

One of the major limitation of this research is finding a group with complete knowledge of the three crowdsourced applications described in Chapter 3. This research will concentrate on evaluating the enviroCar case study. The selected group consists of 17 persons currently working on enviroCar data. The other major problem is the current state of knowledge on the use of Space-Time Cube plug-in in ILWIS platform. So, in order provide the required knowledge on the application, the participants must be interacted with the STC environment. But, due to lack of time and resources the interaction of STC with the participants is not possible. This will affect the understanding of the users while presenting the results.

Considering the available time and available resources, in this research the questionnaires method is thought to be appropriate to conduct usability analysis. Using questionnaire method, both the subjective data and the objective data (correctness of answers) can be acquired. Further, this will help in providing the information on two commonly used measures of usability (effectiveness and satisfaction). The information regarding the efficiency of the visualization cannot be calculated using this evaluation method. In order to provide the information regarding the efficiency, the user needs to be assigned with tasks to be performed on the visualization platform, which is not possible with large number of user groups and in less time. So this research is confined to test the effectiveness and satisfaction of the participants from this usability evaluation.

6.3. Usability evaluation on selected group and its objective

In order to conduct the usability analysis on the preliminary results of crowdsourced data, which were discussed in the previous chapter, a questionnaire evaluation was held. The usability evaluation took place on 29th January, 2015 for about 30 minutes in IFGI, Munster University. As the participants were not able to interact with the Space-Time Cube (because of some technical reasons), the evaluation was started with a ten minutes presentation (Appendix-3) in which, the complete information regarding the visualization and its applications were demonstrated with some examples. The demonstration is followed by questionnaires secession (See Appendix 1) for 15-20 minutes. The usability evaluation was executed to evaluate the preliminary results and cover one of the sub-objectives of this research.

The objectives of this evaluation are:

1. To measure the user understanding of the enviroCar applications in STC (by opinions).
2. To identify new applications of STC on enviroCar data.
3. To measure how well the STC helps in identifying the spatio-temporal patterns.
4. To identify the user preferences, which might help for further application development.

Based on the objectives the questions were posed to the participants in the questionnaires secession. The questionnaires booklet was divided into three main parts:

1. **First part:** The general information regarding the participant's background is collected in this part. This is important to estimate the accuracy of the answers given by the participant.
2. **Second part:** The second part consists of four questions, designed on the preliminary analysis results achieved in Chapter 5. The results will be displayed, and the questions are asked to retrieve the participant's understanding of the visualization. These questions are designed to retrieve both objective and subjective answers from the participants, to make them more accurate. These

questions are formulated to retrieve metrics on the “effectiveness” of the visual representation. From their responses, the amount of understanding of the participants will be derived, and their comments will be noted for potential improvement of the visual representation.

3. **Third part:** Unlike the second part, the third part consists of four subjective questions. From these questions, the user experience with enviroCar data and their impression of the use of Space-Time Cube for spatio-temporal analysis will be questioned. In this part, the questions are asked to retrieve the “satisfaction” of the user with the visual representation. The user comments will be noted and can be used in other related works in the future.

The evaluation group consists of 17 participants in total. All the participants are currently working with enviroCar data in Munster University. Out of 17 participants, 10 participants are doing their M.Sc. (Geo-Informatics), 3 are doing their bachelor's (Geo-Informatics), two working professors, one Ph.D. researcher and one Post. Doctoral researcher. The combination of different participants is also an important part of this evaluation, as it gives diversity of user opinions regarding the visualization environment depending on their level of expertise. For example, it is expected that the professors and Ph.D. researchers will better understand the application and redirect their opinions on the satisfaction and effectiveness of the visualization environment, to explore the spatio-temporal data, when compared to students. This is because of their level of experience on enviroCar data and comprehensive understanding of spatio-temporal pattern analysis.

All the answers given by the participants were noted in the questionnaires booklet (Appendix-1) and stored for the analysis of results. As the participants were not able to interact with the visualization environment, the preliminary analysis results of STC were printed on A3 size paper with enlarged pictures to make the participants more understandable. The results will be described in the following section.

6.4. Evaluation results

The participants give answers both in the subjective and objective form, reflecting their opinion on the visualization. Based on their opinions, results of usability analysis were obtained. These results can help to attract the attention of other similar research to some extent and also help in the development of the application. However, it is important to note that, conducting the questionnaires session with the selected group, not necessarily gathers all the information needed for the evaluating this research. For example, this evaluation is mostly concentrated on evaluating the enviroCar data. Comparing these evaluation results with the other case studies (The nature of these case studies is entirely different) is not completely appropriate. Hence, the results obtained from this evaluation are only used to identify the user understanding on different applications of STC.

In the following section, the results obtained from the first part, second part and the third part questionnaire secession will be discussed. The participants are not directly interacted with the visualization, instead of that they are provided with the snapshots of the results obtained in the previous chapter.

6.4.1. Part-1 evaluation results

As mentioned before, all the questions in this part are formulated to test the user experience and knowledge on geo-visualization and Space-Time Cube. Although all the participants participated in evaluation are related to Geo-Information Science, the level of knowledge will differ. For example the participants who participated in the evaluation are from three different knowledge levels, they are bachelor students, MSc students and researchers (PhD, Post doctoral and Professors). Pie charts in Figure 6.1 represents results of the part-1 evaluation.

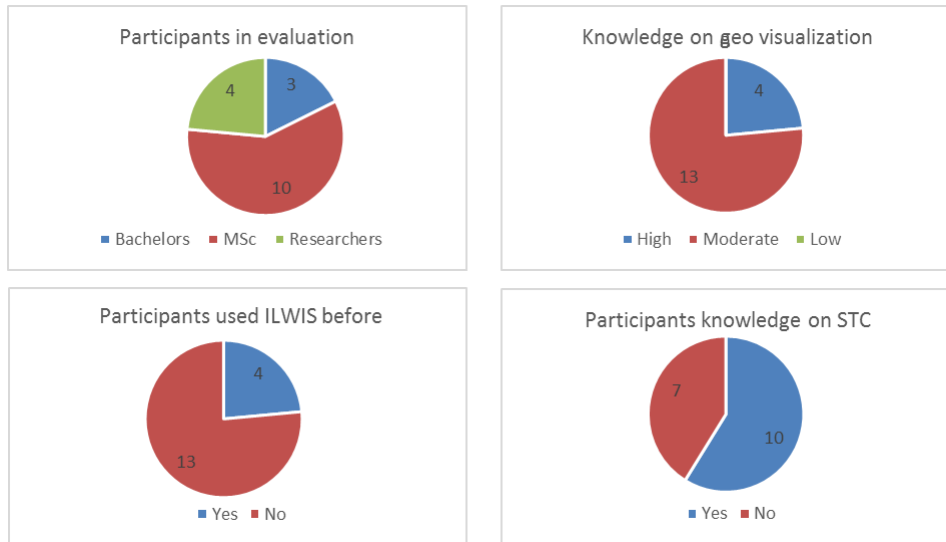


Figure 6.1: Pie charts representing the evaluation results of part-1

6.4.2. Part-2 evaluation results

The questions in this section are formulated to test the effectiveness and understanding of the visual representation. All the questions in this part are both subjective and objective based questions. Based on the objective answers, evaluator can identify the participant's understanding of the representation, and depending on the subjective answer evaluator can further justify the extent of participants understanding. Subjective answers also help in interpreting different views of the participant.

The below image (Figure 6.2) represents the enviroCar data in Space-Time Cube environment. In a Space-Time Cube, the X and Y axis represents the base map (OpenStreetMap), and the Z axis represents the time (2013-2014). As the enviroCar data is a continuous data, it is represented with the help of Space-Time Path (the individual points are connected in the order of time). As you can see the car tracks are irregular in shape, the size and shape of the track represents the emission of CO₂ and the color of the track represents the Car manufacturer. The legend is provided below for reference. Also check the images provided in A3 format for reference.

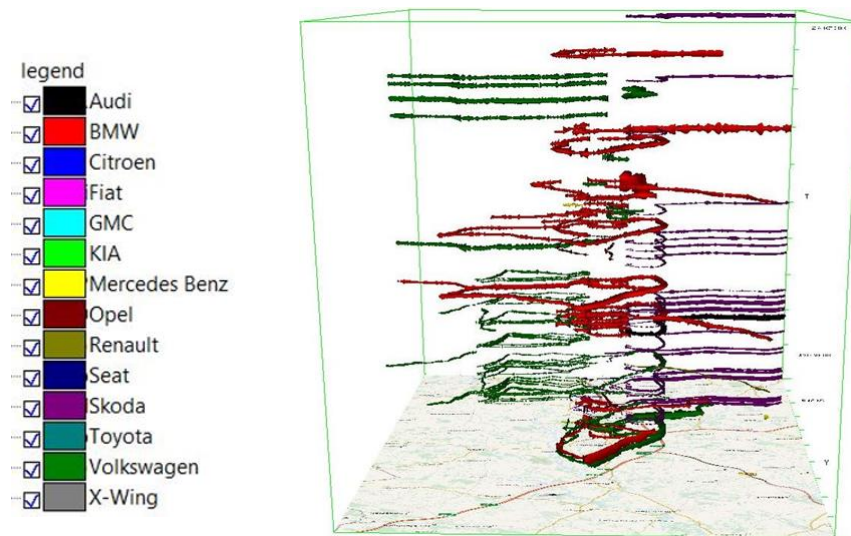


Figure 6.2: Front view of STC

Question 1: In order to answer this question just compare the first 5 images provided in A3 format (Appendix 1). Here, thickness of the track represents the emission of CO₂ and color represents the car manufacturer. Do you think this representation can lead to spatio-temporal pattern analysis of enviroCar data? If so what kind of patterns can you recognize in these images?

This question is formulated primarily to test the user interpretation on the visualization. It proved that all participants understood the representation and gave multiple opinions on the spatio-temporal patterns they identified based on the pictures given to them; the non-interactive nature restricted their exploration capabilities. Still, they identified different patterns like the movement of similar car manufacturers in one particular ways, emission of different CO₂ values for different car manufactures (BMW, AUDI, SKODA, etc.), identification of anomalies in the CO₂ emission and so on. But interestingly there is one participant who commented that this visual representation is not useful for such spatio-temporal pattern analysis on car tracks. According to him the CO₂ emission varies for different cars depending on the size of the car, which is not represented in the visualization. So he interpreted that this visualization might lead to misunderstanding of the information available. This issue will be explained in the following discussion section.

Question 2: Space-Time Cube has the zoom-in, zoom-out and rotate options through which we can manipulate the cube. As a result of these features, the image below (Figure 6.3) shows the anomalies identified on one side of the cube with high CO₂ emission values. So, do you think it is helpful in identifying the anomalies in the data? If so what kind of anomalies can you identify?

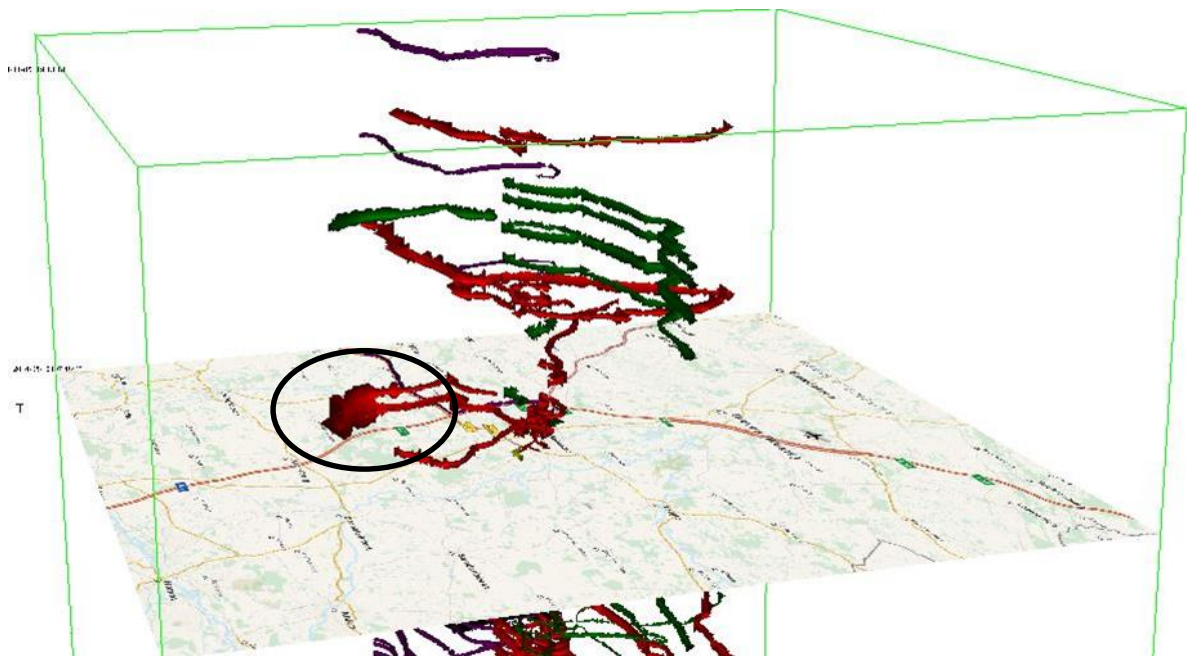


Figure 6.3: Finding anomalies in STC

This question is formulated to test the user understanding with different tools in STC, like the zoom-in, zoom-out and panning of the cube. These tools also play an important role in spatio-temporal pattern analysis. The participants were shown an identified anomaly of CO₂ emission and asked whether this is helpful in identifying such anomalies further.

Every participant understood the representation and gave multiple interpretations. Most of the participants accepted that this representation is helpful in identifying anomalies. However, there are some participants (three) who think other visualization methods would have more advantages in finding anomalies compared to the visualization shown in this question. According to their interpretation the representation

shown in this question is helpful in identifying the outlying values of CO₂, but it will be hard to analyze anomalies when there is a large amount of data. So, instead of this representation they preferred statistical analysis or time series analysis for this visualization. Figure 6.5 represents the comparison of objective answers given by the participants on the identification of anomalies in enviroCar data within STC environment.

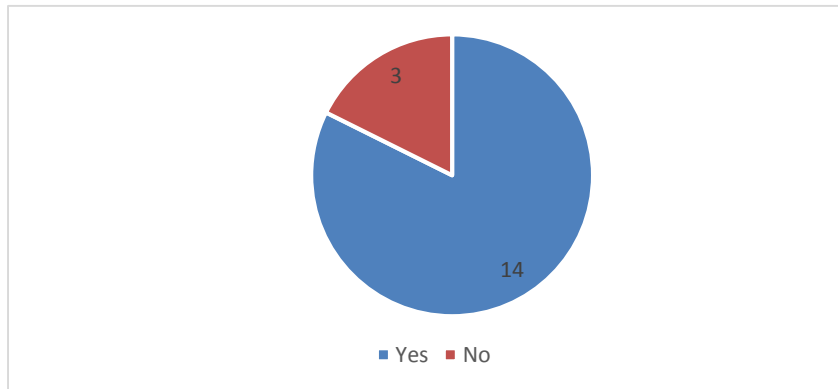


Figure 6.5: User answer on the identification of anomalies using STC

Question 3: The below image (figure 6.4) also represents CO₂ emission along the tracks. On further spatio-temporal analysis on one particular track, we identified that the CO₂ emission is being reduced when the car is taking the turns on the roads or when the speed of the car is reduced. When the result of this track is compared with the rest of the tracks, it showed the same. Do you think this kind of analysis is helpful for spatio temporal analysis?

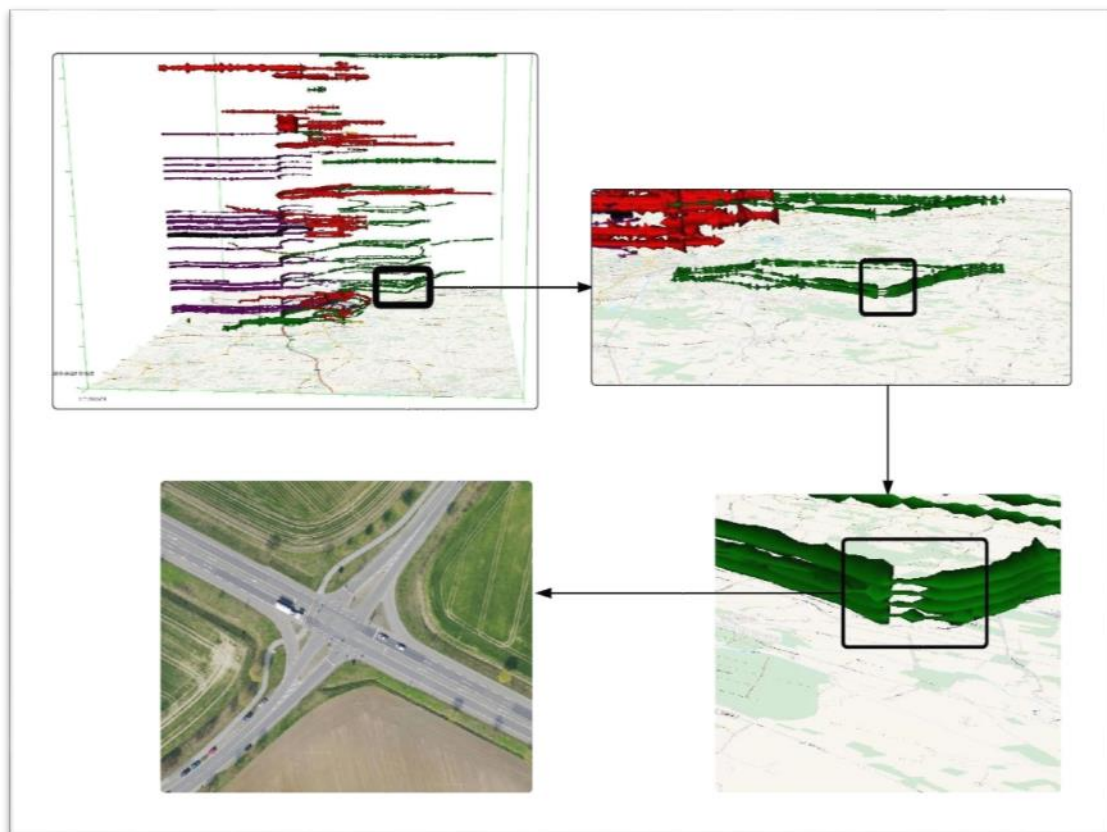


Figure 6.4: Spatio-Temporal Analysis in STC

This question is formulated to test the user understanding of the spatio-temporal analysis conducted within Space-Time Cube environment. The analysis conducted in this question is based on the three fundamental components (when, where and what) of Triad framework. The “when” component represents the time information, “where” component represents the location information and “what” component represents the attribute information, such as the emission of CO₂ and speed. Combining these three components of the triad framework a fundamental question is formed (where + when → what) and the analysis is conducted (explained in the previous chapter).

Every participant understood the representation and gave multiple interpretations on the visualization. Participants agreed that this visualization is helpful for such spatio-temporal analysis and gave their comments on it. Most of them thinks that this kind of analysis helps in understanding the cyclic traffic patterns. However, there are both positive and negative answers to this question. As the result obtained in this analysis is an expected outcome, some participants are more interested in identifying the other reasons for such decreased CO₂ values. This interpretation of the participants will be further discussed in the next section. Figure 6.6 represents the comparison of objective answers given by the participants in the identification of spatio-temporal patterns of enviroCar data within STC environment.

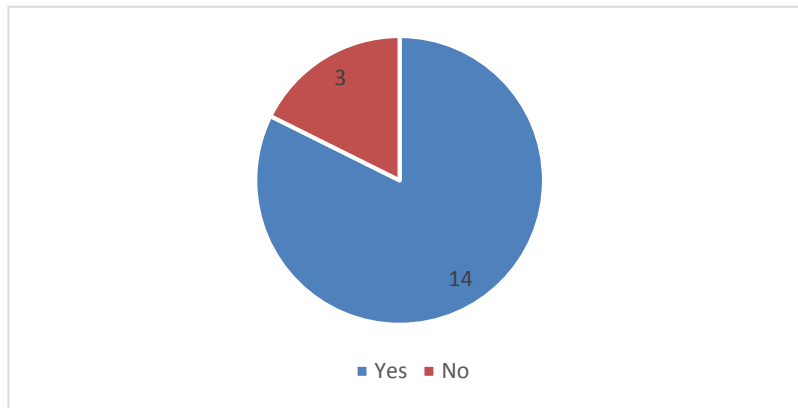


Figure 6.6: User answer to the identification of spatio-temporal patterns in STC

Question 4: The image below represents the posts made by the user at the time of Sandy Hurricane disaster in United States region. The posts at the time of disaster and ten days after the disaster were collected to understand the user participation. The image-a represents the posts made by the user (different colors represents different time of the day), image-b represents the sandy hurricane track (color and size of the track represents the wind pressure) and image-c represents the combination of both layers from image-a, & image-b. You can also compare the images-6 and image-7 in A3 format (Appendix 2) for reference. We came up with an idea of understanding the spatio-temporal patterns of human behavior by studying movement of the Sandy Hurricane track along with the posts made by the user on Flickr. Do you think this method is useful in understanding the spatio-temporal patterns? (Legend followed for this images were given below)

This question is formulated to test the user understanding in different applications of crowdsourced data, other than enviroCar in STC. As described in Chapter -2 the STC plays an important role in spatio-temporal pattern analysis of crowdsourced data. For this question, the data gathered from Flickr case study is implemented within STC. The different colors of the stations (Flickr posts) represent the different periods (morning, afternoon & night) of a day. The space-time path represents the Sandy hurricane movement. Both the stations and the path are implemented into the Space-Time Cube at the same time to compare the human behavior at the time of disaster (explained in the previous chapter).

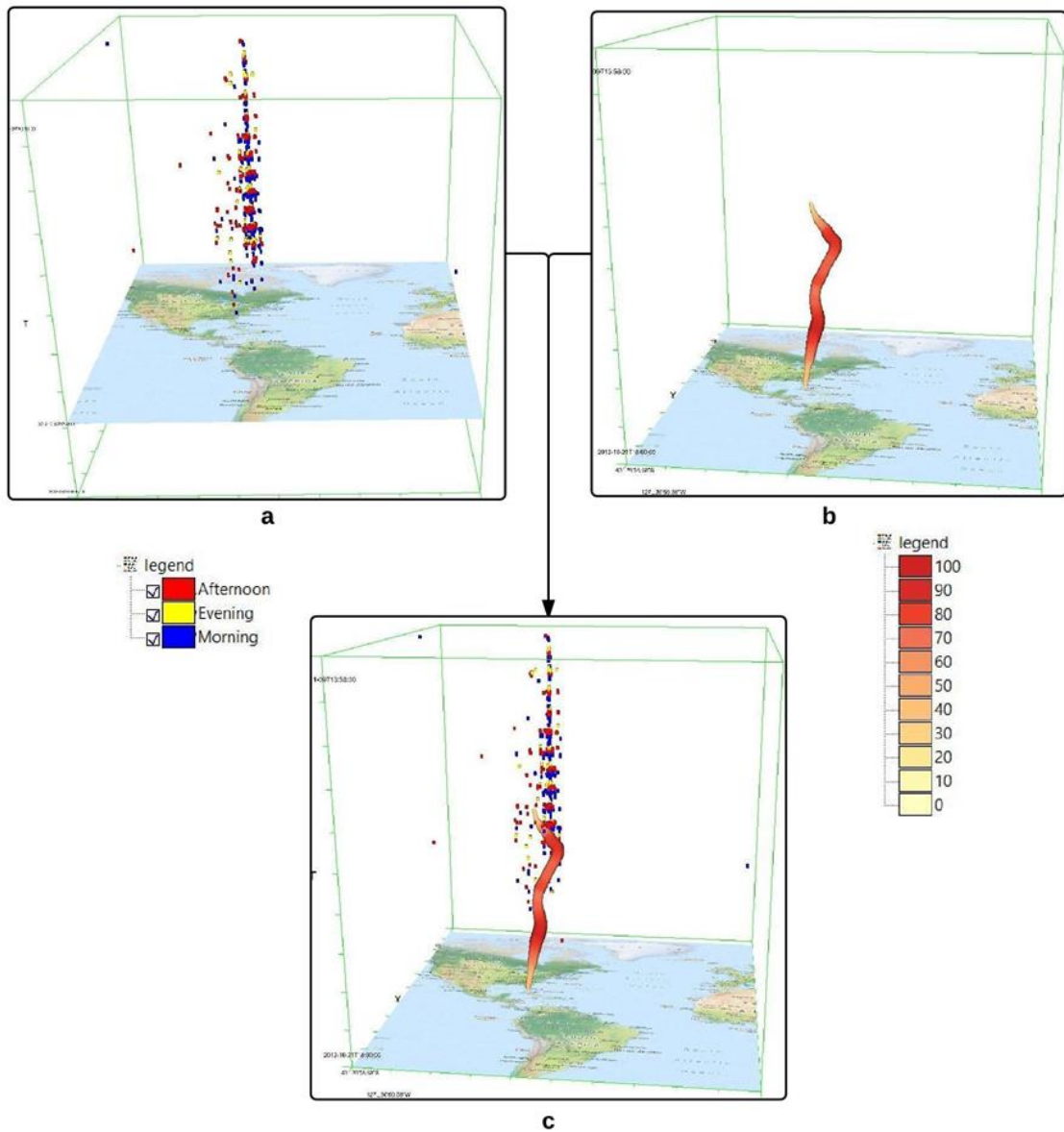


Figure 6.7: Flickr data analysis

This question is a bit complicated question when compared to the rest of the questions in this evaluation, as two types of data (point data and continuous line data) are compared parallel to each other, which made the participants a bit confused. There were multiple interpretations of this question, most of the participants accepted this visualization as a helpful visualization for the better understanding of the human behaviors. But there are some negative remarks along with the positive ones. Some participants interpreted that there is no high correlation between the movement of the Hurricane path and the posts made by the Flickr users. Mostly they interpreted that the visualization is helpful, but the representation of the Flickr posts with different colors (morning, afternoon & night) is not quite understandable. This interpretation of the participants will be further discussed in the following section. Figure 6.8 represents the comparison of objective answers given by the participants in the identification of spatio-temporal patterns in Flickr data within STC environment.

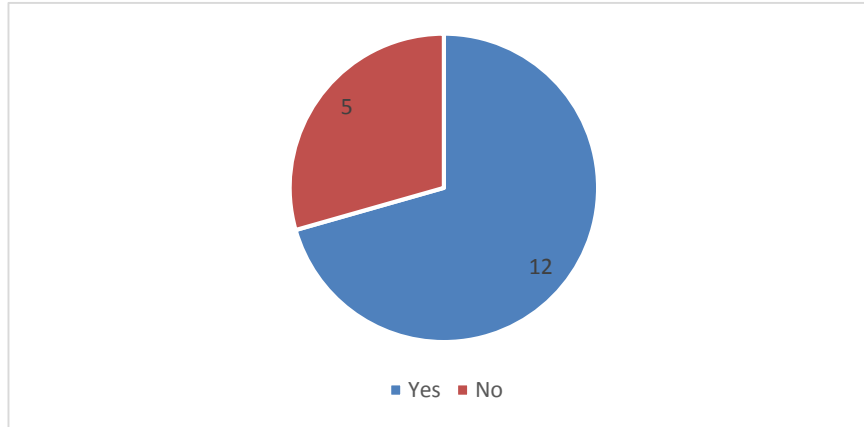


Figure 6.8: Answer to the identification of spatio-temporal patterns of Flickr data using STC

6.4.3. Part-3 evaluation results

The questions in this section are to identify the new applications of enviroCar data and to test the satisfaction of the user on the visual representation. All the questions in this part are subjective.

Question 1: How do you intend to use enviroCar data for analysis?

This question is intended to understand the participant's experience with the enviroCar data. It will also help in identifying some new visualizations that can be performed with the assistance of STC on enviroCar data. In the past the Space-Time Cube is mainly used as a tool to visualize movement data, but by identifying the different visualizations of the enviroCar data, might lead to new research approaches in future.

As the usability evaluation was conducted on the selected group that is currently working on the enviroCar data, this question had a lot of responses. The participants are working on different projects like the hotspot analysis, identification of anomalies, detection of emergency brakes using the GPS values, comparison of the enviroCar data with the local bus data to check the convenience of the passengers during rush hours, traffic management and so on. Some of these working projects can also be implemented in STC for analysis.

Question 2: What other applications do you foresee with the help of Space-Time Cube?

This question is asked, to obtain the participants view on the use of Space-Time Cube in different applications. By introducing complete new application of STC (Flickr data analysis) in question four of Part-2, the participants were given an overview of different applications in STC. This might help in getting some new ideas on this visualization and also assist in identifying new research areas for the future.

There are many vague responses from the participants because of being an open question. They thought of applying this visualization in the study of flood management, identification of spread of diseases, analyzing social media responses, comparison of animations maps, traffic planning management and applying 3D view to get more insight into the data. There is no coherent response from two participants for this question. There are some participants who did not respond to this question, this might be because of the less response time (15-20 minutes to complete the evaluation) given to answer these questions.

Question 3: If possible how do you intend to use the Space-Time Cube in enviroCar data analysis?

This question is asked, to obtain the participants view on the use of Space-Time Cube in enviroCar data analysis and also to identify the different applications of STC on the enviroCar data. The analysis

conducted on the enviroCar data for this research is represented in the previous chapter, but the results that can be achieved using the STC on enviroCar data is beyond the scope of one user interpretation.

This question is also an open question to the participants. Most of the participants answered this question by interlinking their working project on enviroCar data with STC. So, the result achieved in for this question is mostly the same as the first question. Some participants hesitated to answer this question, this might be because of lack of time to interpret their answer in the short period of time.

Question 4: Do you think Space-Time Cube is a useful tool for Spatio-temporal analysis?

This question is an open question to identify the participants' opinion on the use of Space-Time Cube in spatio-temporal analysis. The participant's understanding is asked basing on the overall demonstration and questionnaires secession. Depending on their response, the application of Space-Time Cube in the use of the spatio-temporal analysis can be rated.

As expected all the participants were impressed with the applications of Space-Time Cube and gave a positive response to this question. Everyone accepted that this visualization is helpful in the identification of spatio-temporal patterns. However, along with the positive comments, they also mentioned some negative effects of this visualization as well. They thought that, STC is very helpful in the overall analysis of spatio-temporal patterns of the data, but it is weak for statistical analysis. They also mentioned that the usefulness of STC is very limited to some application, especially to spatio-temporal analysis.

6.5. Discussion of the evaluation results

Both positive and negative comments were gathered from the usability evaluation, conducted on the selected group. From the overall outcome obtained from the participants, STC seems to possess the right characteristic to perform spatio-temporal analysis on the crowdsourced information. It is important to note that, only a subset of one case study data (enviroCar case study), which is implemented in the previous chapter is demonstrated to the participants and used for the evaluation. However, the comments received from participants are helpful in obtaining the overall outcome.

This usability evaluation is mostly emphasized in making the participant understand the visualization and gather their views. However, it is hard to present all the feedback received on the visualization. So the most significant issues are summarized below:

1. In question one of part-2, the participants pointed out that the emission of CO₂ depends on the size of the car, but the representation also visualizes the different car manufacturers with different colors. Depending on the car manufacturer the emission of the CO₂ will vary to some extent, as the high speed cars like BMW, BENZ and AUDI will emit high CO₂ values, through which we can identify the difference in the emission of CO₂. This is one solution, but there are other visual solutions for this as well.
2. In question two of part-2, the participants pointed out the issues in the identification of anomalies with large data. The increase in the data leads to the cluttering of the data, which indirectly leads to the missing of information. But, the use of color variable plays a crucial role in the identification of anomalies. Instead of using the size to represent the CO₂ values, the use of color can help in the identification of anomalies. It is the matter of changing the visual variable that represents data.
3. In question three of part-2, the participants pointed out that the result obtained from this analysis is the obvious, as it is a known fact that the CO₂ emission from cars will be high at the traffic signals. But, the point to be noted here is, it is an example shown to explain the kind of spatio-temporal analysis that is conducted within STC environment. Such analysis can be applied for the identification of speed bumps as well.
4. In question four of part-2, the negative comments given by the participants is due to the lack of interaction with the STC. The participants need to interact with the STC for understanding such complicated visualization. However, they got confused in identifying the data integration between

the Flickr posts and the Sandy hurricane track. To find the data integration between the posts and the hurricane, we need to use the movable base map. The hurricane movement can be tracked by moving the base map with respect to time and parallel to that the posts made by the user will start to appear on the map. As the time width of temporal scale representation (z-axis of STC) for this visualization is 20 days, the identification of high user participation in the span of a day is harder. So the data is further divided into different categories of a day (morning, afternoon and evening).

5. In part-3, the participants interpreted that the visualization environment looks attractive, but the visualization of large data sets might be complicated. However, the overall opinion of the visualization is positive, and the participants admitted that the visualization is helpful for spatio-temporal analysis.

After conducting the usability evaluation, the objectives mentioned at the beginning of this chapter were achieved by analyzing the responses of the participants. The participants understanding on the enviroCar application in STC environment is measured by analyzing the answers gathered from the part-2 evaluation. From the subjective answers given by the participants for most of the questions is used in identifying new applications of STC on enviroCar data. The examples provided in part-2 evaluation on the possible spatio-temporal analysis that can be conducted on STC, helped in collecting the responses of the participants, on the usefulness of the STC in spatio-temporal pattern analysis (part-3 evaluation). Also, the open questions in the part-3 evaluation helped in identifying the user preferences, which might help for further application development.

6.6. Summary

This chapter described the results of the usability analysis undertaken on the preliminary analysis results of the enviroCar data. The test was undertaken by using the questionnaires method with the help of a group consisting of 16 participants. The users were asked to give both subjective and objective answers, depending on their knowledge of enviroCar application. Based on these subjective and objective questions the effectiveness and satisfaction of the visualization environment are analyzed. This evaluation is conducted in three parts. In the first part, questions related to the background of the participant is questioned. In the second part, the preliminary analysis results of my research will be displayed, and the questions are asked to test the participant's understanding (effectiveness is tested). In the third part, the user experience with enviroCar data and their impression of the use of Space-Time Cube will be questioned (satisfaction on the environment is tested). The overall results show that the representation of enviroCar data in Space-Time Cube had its advantages and disadvantages. Most of the disadvantages is due to the problems in understanding the Space-Time Cube. Regarding the comments given by the participants on the visualization, can be used for other related works in the future and make it a more robust application for data exploration.

7. CONCLUSION AND FURTHER RECOMMENDATION

7.1. Introduction

In the previous chapter, the evaluation method selected and the results of the evaluation process were discussed. This chapter answers the research questions in this research, and provides final remarks on the usage of STC and recommendations for further research.

7.2. Conclusions

7.2.1. Revisiting the research question

The answers to each of the research questions proposed in Chapter 1 can be summarized as follows:

1. What are the different types of crowdsourced geo-information?

This research question is answered in Section 2.2. The concept of crowdsourcing, and the different kinds of crowdsourced geo-information have been explained in that chapter. There have been multiple interpretations for crowdsourced geo-information. But, it is believed that the most known phenomena are the “Volunteered Geographic Information” (VGI) and “The Citizen Science”. Both phenomena are used to collect and share the geo-information with the involvement of large user groups, but the only difference lies in the participants taking part in the information collection. In Citizen Science projects, the participants are expected to have knowledge of the application. Whereas in the VGI projects, the participants may or may not have any knowledge of the application. Sometimes, in VGI the participants do not even know that their action can lead to geo-information collection (e.g. social media data). Keeping these aspects in mind the case studies were selected in this research. For example, the short event dependency in the Sandy use case, the repetitive character of enviroCar case study, and the attribute richness in the OSM case study.

2. How is crowdsourced data produced, and what are the consequences for its quality?

This research question is also answered in Section 2.2. The crowdsourced data is created by the collaboration and sharing of large groups of people, using web as a resource. The participants collecting and sharing the data may or may not have knowledge of the application. Hence, the answer to this question can be interpreted in two ways. One way is by considering that the data collected from the participants with knowledge of the application domain (Citizen Science), and the other way is by considering that the data is collected from the participants without any knowledge on the application (VGI). There is a possibility of achieving good quality if the data is collected by participants with knowledge on the application. However, the quality of the application must be described according to the “fitness for use” (Heipke, 2010). The volunteers participating in the crowdsourced events are also high in number. Most of the volunteers are motivated to participate in the crowdsourced events to map their surroundings or the places they visited. This is also one of the major factor for increasing the quality of the data. Also with the advancement of technology, the data collection with high accuracy is becoming an easy task with the use of GPS.

3. What are the most commonly used visualization techniques for crowdsourced data analysis?

This research question is answered in Section 2.4. The traditional visualization methods like the Single static maps, series of static maps and animation maps are the most commonly used visualization techniques for crowdsourced data visualization. The static maps are the most common and well-used method for this representation. But, by using the static maps it is hard to represent the temporal changes and the spatial changes at a time. So the single static maps are combined with other geovisualization techniques (e.g. graphs) to represent the data. Series of Static maps and

animations maps are other important visualization techniques to represent the crowdsourced geo-information.

4. Among those visualizations, which are best suited to deal with the character (location, attribute, and time) of crowdsourced data?

This question is answered in Section 2.4. Single static maps are not that suitable for representing the spatial and temporal aspect integrally. Animation maps are much more utilized for representing the spatio-temporal changes, but the information extraction is difficult. So, there is a need for a visualization that have high interactive capability and can deal with spatial and temporal components of the data. Hence, in this research the Space-Time Cube visualization environment which have high interaction capabilities and can clearly deal with both space and time components is introduced to visualize the crowdsourced information. The next research question will answer how STC will fit in for this visualization.

5. How does the STC fit-in?

This question is answered in Section 2.5. The STC is introduced by Hägerstrand at the end of sixties, to study the movement of individuals in day to day activities. The main advantage of the STC is that it has a horizontal plane to deal with the spatial data component and the vertical plane to deal with the temporal data component. By the combination of these two planes (horizontal and vertical) in one visual environment, the Space-Time Cube is designed. This helps in representing the spatio-temporal data at a time in one environment for analysis. On the other hand, the data collected for this research is volunteered data, which consists of spatial component (where), temporal component (when) and the attribute information (what). This is the main reason for selecting the STC as the visual representation of this research. Using these fundamental components of the data, basic questions can be formulated from the triad framework and can be visualized in STC for spatio-temporal analysis.

6. What is the objective of the enviroCar use case, and what is the nature of its data?

This question is answered in Section 3.4. enviroCar is a crowdsourced application in which the volunteers are collecting the data to get insight into sustainable mobility. Any participant can collect this data, but they are supposed to have an OBD-II adapter connected to the car engine and an android mobile device. The knowledge of the application is not necessary for the participants, but they require equipment to collect data. So this case study comes from an intersection of VGI and Citizen Science.

The main objective of this case study is to implement the car track data into STC and check whether it works for spatiotemporal analysis. This case study data is a continuous track data collected from different car manufacturer (BENZE, AUDI, etc.). The repetitive character of enviroCar case study will be studied in the STC to identify the spatio-temporal changes caused by the change in time and attribute.

7. What is the objective of the Flickr use case, and what is the nature of its data?

This question is answered in Section 3.3. The Flickr is a social networking application in which the users can share their pictures with their friends. The users posting the pictures in this application do not post them with a purpose, and they do not know how this data can be utilized further. But the data analysts collect this data for studying the human behavior. So this application falls under the category of VGI.

The Sandy Hurricane disaster is selected for analysis in this case study. The main objective of this case study is to visualize the posts of Sandy Hurricane disaster of United States region in STC and

check whether it works to identify the spatio-temporal patterns. The short event dependency in the Sandy use case is tested in the STC environment by comparing the Flickr posts with the Sandy hurricane track data. In this way crowdsourcing can be helpful in studying the disaster events such as hurricanes and STC can help in visualization and spatio-temporal exploration of the data.

8. What is the objective of the OSM use case, and what is the nature of its data?

This question is answered in Section 3.5. The OpenStreetMap is a crowdsourced application, created with an objective of creating a free and editable world map for all the users. The main objective of this case study is to analyze the OSM mappers' participation under different circumstances in different regions of the world. For achieving this objective, Miyagi region of Japan (Tohoku earthquake), Haiti region (Haiti earthquake), Overijssel region (developed country), and Tanzania region (developing country) are selected for analysis. The OSM data is with rich attribute information and is hard to visualize. STC plays a prominent role in visualizing such large information by visualizing the data spatio-temporally and dividing the attribute information into different colors. On the other hand, the OSM mappers are supposed to possess knowledge of the application to collect and digitize the geo information. So, this case study falls under the Citizen Science type.

9. How can the data of each case study be implemented into the STC?

This question is explained by combining different chapters (Chapter 2, Chapter 3, Chapter 4 and Chapter 5). The objective of the three case studies is identified in Chapter 3, and the data is collected accordingly. The data collected is spatio-temporal data, consisting of the time information (when) and the spatial information (what) for every point along with the attribute information (what) such as the CO2 emission and speed of the car. Once all these three fundamental components (when, what and where) of the data are identified, the data can be implemented into the STC for spatio-temporal analysis using the Triad framework.

The major limitations encountered during the implementation of data into STC is the identification of time filed by STC. STC had its own time field and it will not identify the other forms of time. So, when the data is imported into ILWIS we need to specify a new time field that is recognized by STC.

10. How to evaluate the results?

This question is answered in Question 12.

11. How to execute a questionnaires session?

This question is answered in Question 12.

12. How was the solution experienced by the user of the case study?

The 10th, 11th, and 12th questions are all together answered in Chapter 6. It is hard to find the user group with knowledge of all three crowdsourced applications selected in this research. So, the enviroCar case study is evaluated with a user group that is currently working on that data. The questionnaires evaluation method is thought to be appropriate for this evaluation by considering the time and resources as major factors. The evaluation is conducted to test the participants' effectiveness and satisfaction. Due to the lack of time and facilities, the interaction with the visualization environment is not provided to the participants. This might cause misunderstanding to the participants. To solve this issue a 10 minutes presentation (explaining the concepts of STC and preliminary results) was given to the participants at the beginning of the evaluation and followed by the questionnaires session. There are both positive and negative comments gathered from the

evaluation. From the overall outcome obtained from the participants, STC seems to possess the characteristic to perform spatio-temporal analysis on the crowdsourced information. Most of the participants liked the visualization and also gave their view on how they might use this visualization in the future. However, there are some negative comments as well. The negative comments are mostly on the clutters formed by visualizing the large datasets.

However, it is important to note that, conducting evaluation with the selected group, not necessarily gathers all the information needed for the evaluating this research. For example, this evaluation is mostly concentrated on evaluating the enviroCar data. Comparing these evaluation results with the other case studies (The nature of these case studies is entirely different) is not completely appropriate. Hence, the results obtained from this evaluation are only used to identify the user understanding on different applications of STC.

7.2.2. Advantages and disadvantages of STC

During the processes of achieving the objective of this research, some advantages and disadvantages of the cube were identified. Even, the overall evaluation results shows that the cube had its advantages and disadvantages. So, some of the important factors that highly influence the STC are mentioned below.

Advantages of STC

1. The STC provides a possibility to explore the entire spatio-temporal data at a time.
2. With the advancement in technology, STC is providing the option to visualize, retrieve and explore the spatio-temporal information.
3. The STC includes different concepts to visualize the continuous moving data (Space-Time Path), and individual points (stations) spatio-temporally.
4. Use of different visual variables to represent different attribute information helps in the extraction of spatio-temporal patterns.
5. The use of time bar to move the base map helps in the spatial exploration.
6. Cube manipulation tools (panning tools, zoom-in, and zoom-out tools) play a crucial role in spatio-temporal pattern exploration.

Disadvantages of STC

1. Unskilled users do not readily understand visualization environment.
2. The user needs to have good knowledge of the concepts like the Space-Time Paths and Stations in order to understand the visualization.
3. The zoom-in and zoom-out options available in the STC are restricted to the overall scale of the cube.
4. The functions (querying of data, data selection options & statistical analysis) in the STC still needs development.
5. Visualization of large datasets leads to visual cluttering of data.
6. The STC is limited to spatio-temporal point data analysis.

7.3. Recommendations for further research

During the process of achieving the objective of this research, a number of issues were encountered which could lead to some recommendations for future research. During the course of this research, the data was implemented into the Space-Time Cube for Spatio-temporal analysis and then the usability analysis was performed. From the evaluation results, a number of new opportunities of STC in enviroCar case study were identified, but the time is not sufficient to implement them by the end. Instead of this process, it would

be recommended to perform two usability analyses. One preliminary usability analysis to identify the research that is being carried out on the crowdsourced data and the other usability analysis to evaluate the results achieved by STC. This process would save a lot of time in identifying the objective of the case study.

The STC is a complicated visualization environment that is not easily understood by the users. In order to make the users understand the actual notation of STC, the participants must be provided with an option to interact with the STC during evaluation. On an overall analysis, the visualization got positive responses, but it is recommended to perform an evaluation by allowing the user to interact with the visualization environment.

The objective and the nature of the three case studies is different in this research. On time being, only one case study data was evaluated to estimate the usefulness of the STC, but it would be recommended to evaluate all the three case studies with different user groups to identify the usefulness of STC.

The functions in STC can still be expanded and improved. The zoom-in and zoom-out options of STC are restricting the exploration capabilities of the users. If there is an option to zoom into the data spatially and temporally, it would solve the problems of visual clutters, leading to the identification of new spatio-temporal problems. Also, the select by attribute option is only limited to Space-Time Paths. It will be nice if this option is available for point data selection as well. These options are highly recommended for further developments in STC.

In this research only OpenStreetMap is used as the basemap for spatio-temporal analysis. But, in the STC, there is an option to add different basemaps for analysis. The addition of a thematic map such as the population density map or the animation map for spatio-temporal analysis will also be an interesting point of research in the future.

APPENDIX 1: QUESTIONNAIRE EVALUATION

Usability Analysis

Name :
Email :
Sex :
University :

Dear Munster University Faculty Members,

I am Geo-Informatics MSc student from ITC, Netherlands. I am working on Crowd-sourced geo information and their visualization under the supervision of Prof. M. J. Kraak and Prof. Rob Lemmens. As a part of my research, I am doing a usability analysis on the usefulness of Space-Time Cube in order to evaluate my intermediate results and discover new features. If you decide to participate in this test, you will be shown some intermediate results on enviroCar data (enviroCar tracks), and you will be asked to answer few questions basing on your knowledge of enviroCar data. Do not worry, the questions are not designed to test your intelligence, everything will be explained beforehand.

Aim of the research is to evaluate the usefulness of Space-Time Cube in context of crowd-sourced geo information. As a part of the crowd-sourced geo information, enviroCar data is selected as one of the case studies of this research. For this reason, a fundamental idea of the enviroCar data is required to answer the following questions.

If you are willing to participate, I would like to introduce you to the usability analysis. This evaluation is conducted in three parts. In the first part, questions related to the background of the participant is asked. In the second part, the preliminary analysis results of my research will be displayed, and the questions are asked to test the participant's understanding. In the third part, the user experience with enviroCar data and their impression of the use of Space-Time Cube will be questioned.

This usability analysis might take 10-15 minutes. The results of this analysis and your identity will not be revealed, and you will remain anonymous.

Thank you in advance for your Time and Consideration.

With Regards,
Jothirnadh Sai Guthula

Part-1

1. What is your professional background (study /work)?
 - a. Geo-Informatics (GIS/Remote Sensing)
 - b. Computer Science
 - c. Civil
 - d. Other (Please specify) _____

2. Current Position?
 - a. Ph.D.
 - b. MSc
 - c. Bachelor
 - d. Post. Doctorial
 - e. Professor/Lecturer

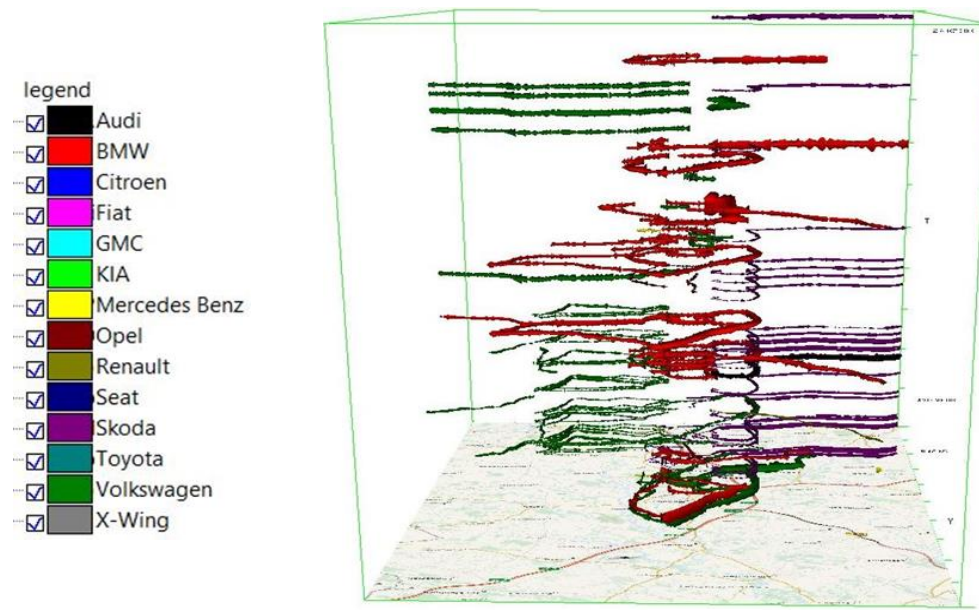
3. What is your level of knowledge about geo visualization?
 - a. High
 - b. Moderate
 - c. Low

4. Have you ever used ILWIS software?
 - a. Yes
 - b. No

5. Have you ever interacted with Space-Time Cube?
 - a. Yes
 - b. No

Part-2

The below images represents the enviroCar data in Space-Time Cube environment. In a Space-Time Cube, the X and Y axis represents the base map (OpenStreetMap), and the Z axis represents the time (2013-2014). As the enviroCar data is a continuous data, it is represented with the help of Space-Time Path (the individual points are connected in the order of time). As you can see the car tracks are irregular in shape, the size and shape of the track represents the emission of CO₂ and the color of the track represents the Car manufacturer. The legend is provided below for reference. Also check the images provided in A3 format for reference.



Front View of STC

1. In order to answer this question just compare the first 5 images provided in A3 format. Here, thickness of the track represents the emission of CO₂ and color represents the car manufacturer. Do you think this representation can lead to spatio-temporal pattern analysis of enviroCar data? If so what kind of patterns can you recognize in these images?

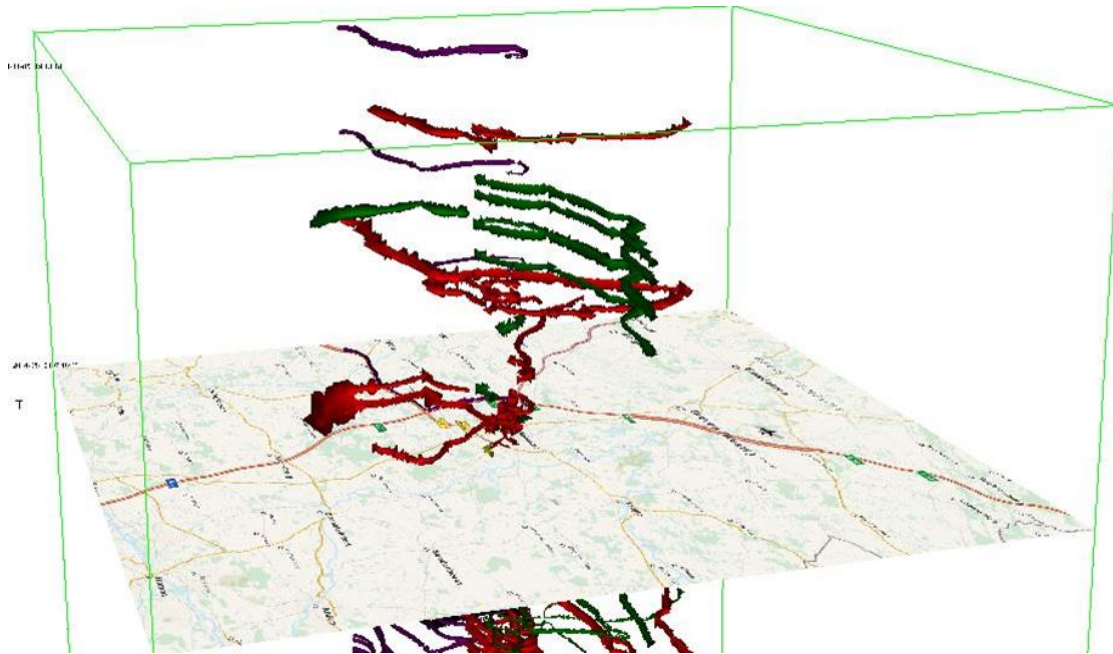
- a. Yes
- b. No

Explain your answer with short notes

2. Space-Time Cube has the zoom-in, zoom-out and rotate options through which we can manipulate the cube. As a result of these features, the image below shows the anomalies identified on one side of the cube with high CO₂ emission values. So, do you think it is helpful in identifying the anomalies in the data? If so what kind of anomalies can you identify?

- a. Yes
- b. No

Explain your answer with short notes

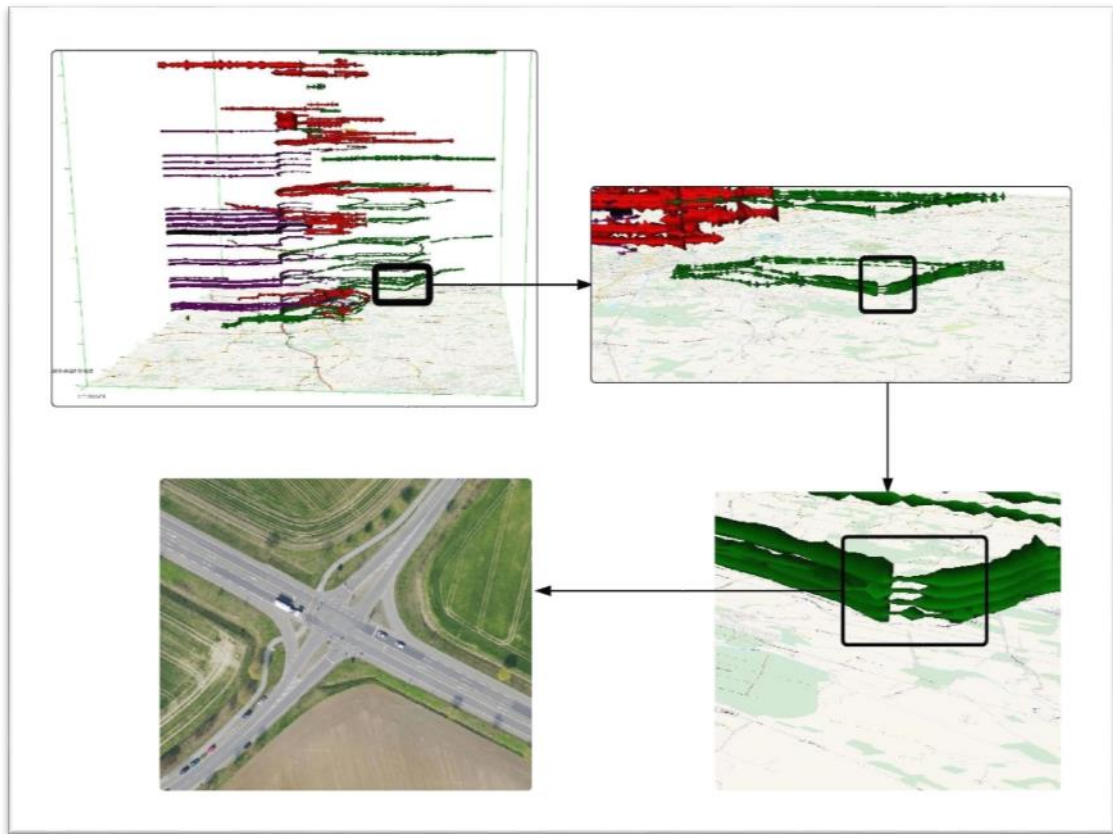


Finding anomalies

3. The below image also represents CO₂ emission along the tracks. On further spatio-temporal analysis on one particular track, we identified that the CO₂ emission is being reduced when the car is taking the turns on the roads or when the speed of the car is reduced. When the result of this track is compared with the rest of the tracks, it showed the same. Do you think this kind of analysis is helpful for spatio temporal analysis?

- a. Yes
b. No

Explain your answer with short notes

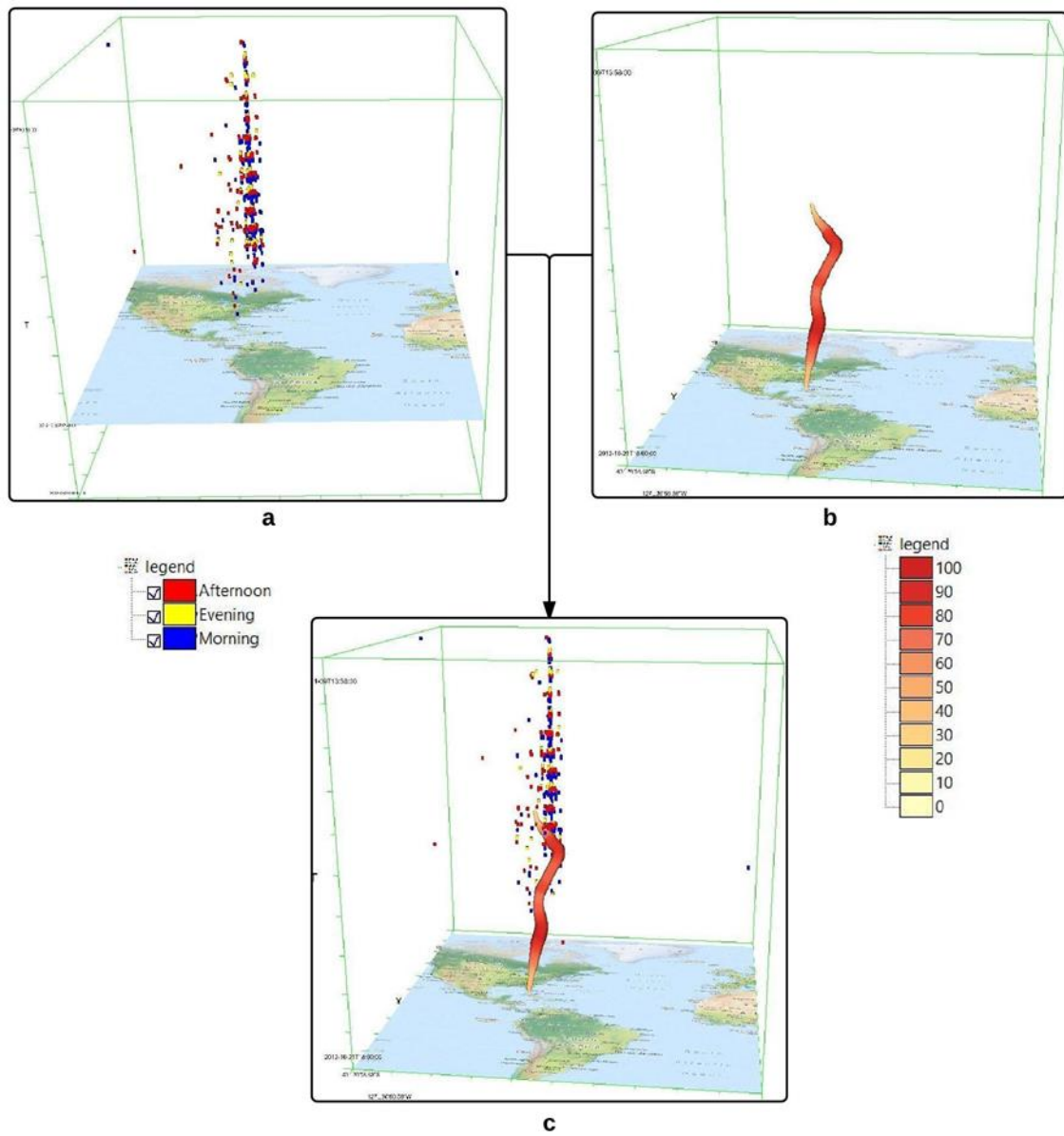


Spatio-Temporal Analysis

4. The image below represents the posts made by the user at the time of Sandy Hurricane disaster in United States region. The posts at the time of disaster and ten days after the disaster were collected to understand the user participation. The image-a represents the posts made by the user (different colors represents different time of the day), image-b represents the sandy hurricane track (color and size of the track represents the wind pressure) and image-c represents the combination of both layers from image-a, & image-b. You can also compare the images-6 and image-7 in A3 format for reference. We came up with an idea of understanding the spatio-temporal patterns of human behavior by studying movement of the Sandy Hurricane track along with the posts made by the user on Flickr. Do you think this method is useful in understanding the spatio-temporal patterns? (Legend followed for this images were given below)

- a. Yes
- b. No

Explain your answer with a short notes



Flickr data analysis

Part-3

1. How do you intend to use enviroCar data for analysis?

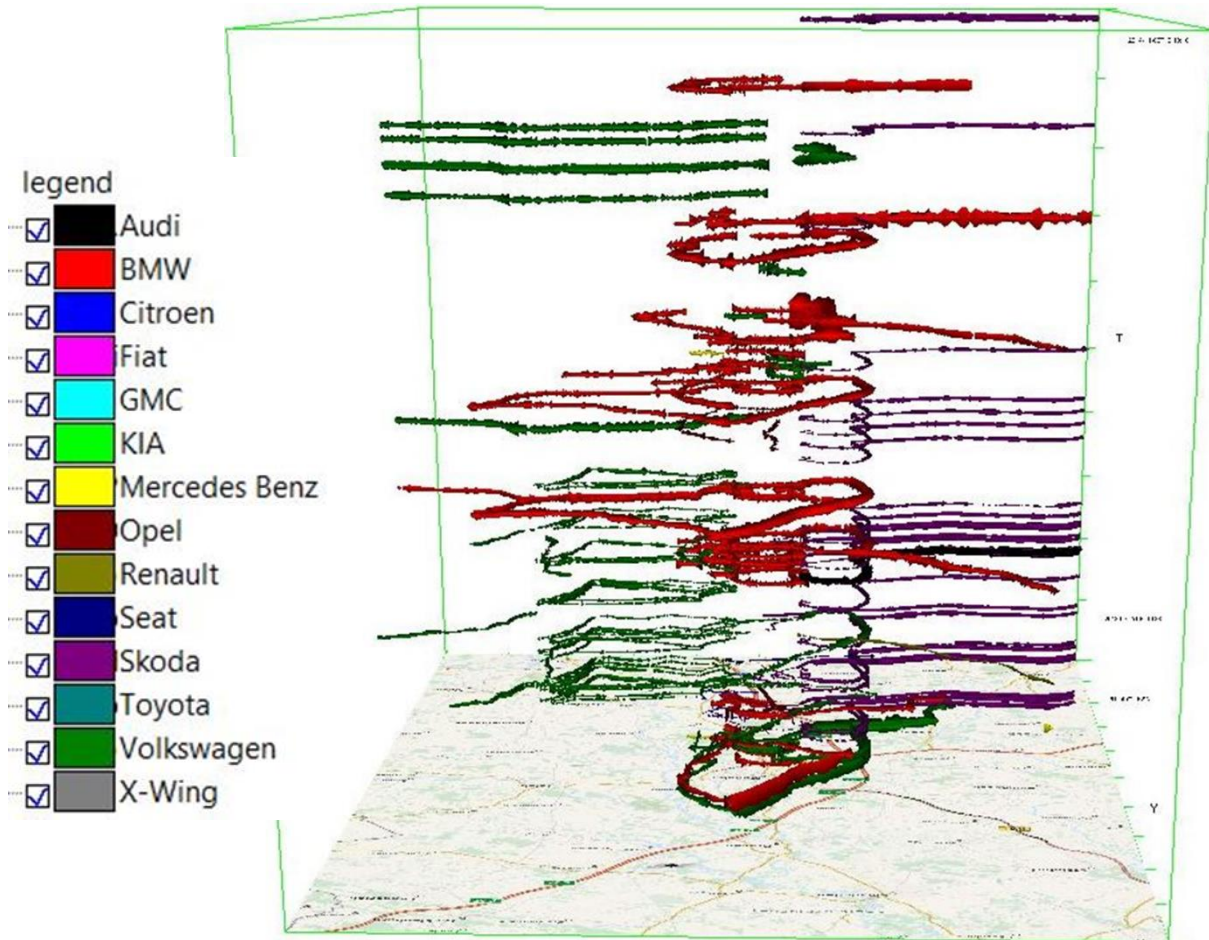
2. What other applications do you foresee with the help of Space-Time Cube???

3. If possible how do you intend to use the Space-Time Cube in enviroCar data analysis?

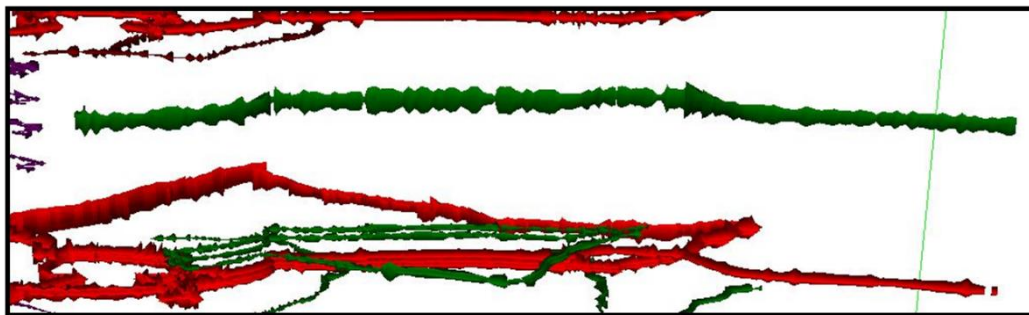
4. Do you think Space-Time Cube is a useful tool for Spatio-temporal analysis?

APPENDIX 2: VISUALIZATIONS PRINTED IN A3

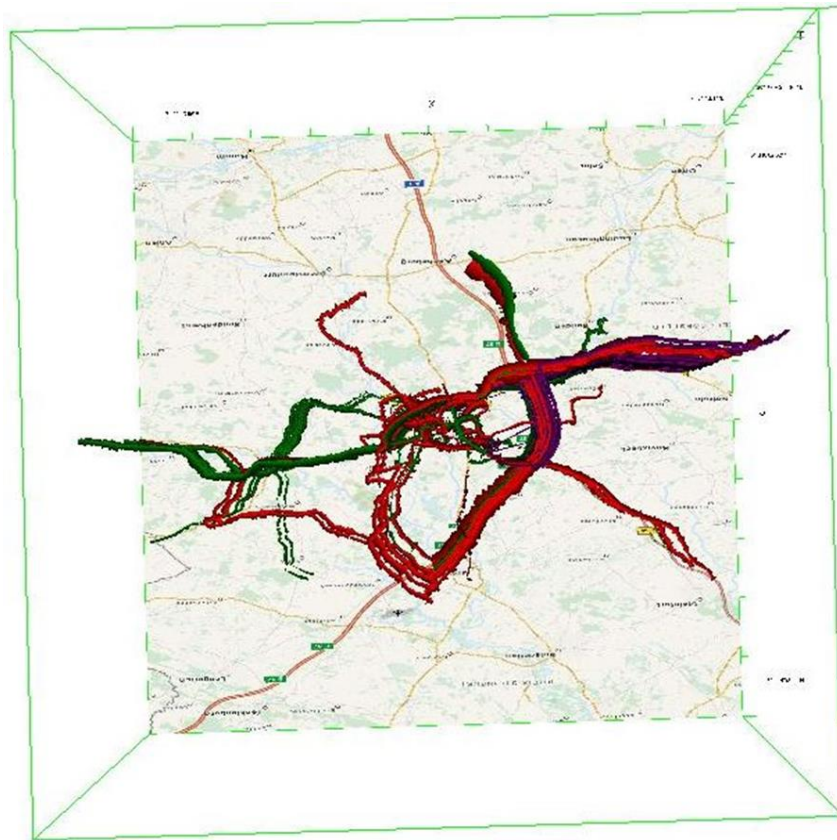
The images used in the questions were printed in A3 format for reference.



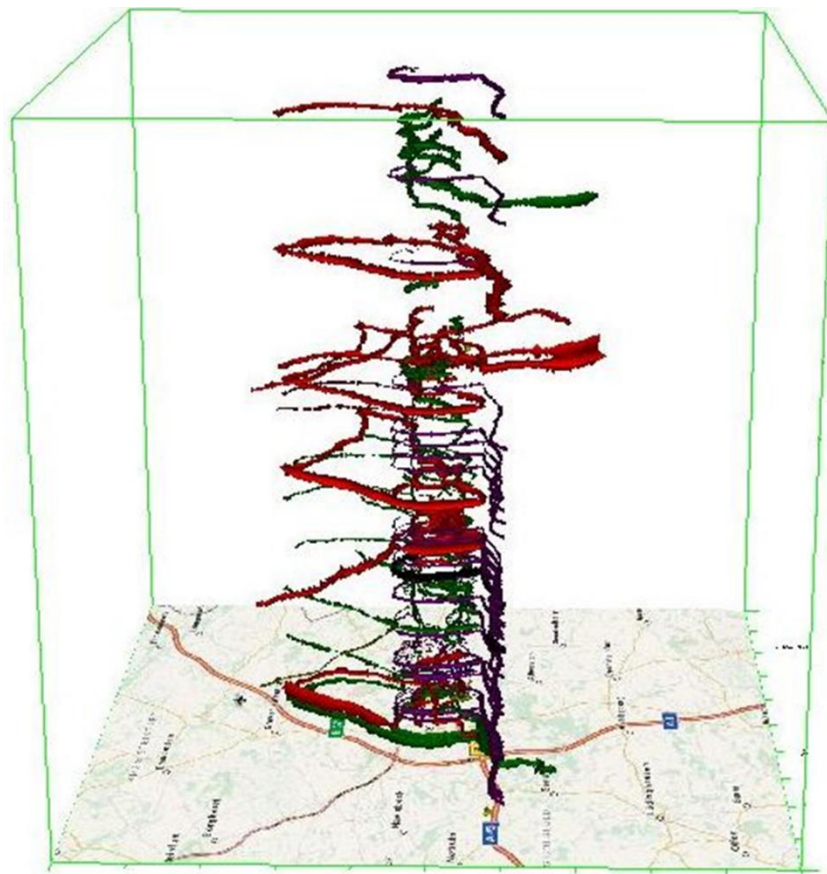
Front View of STC



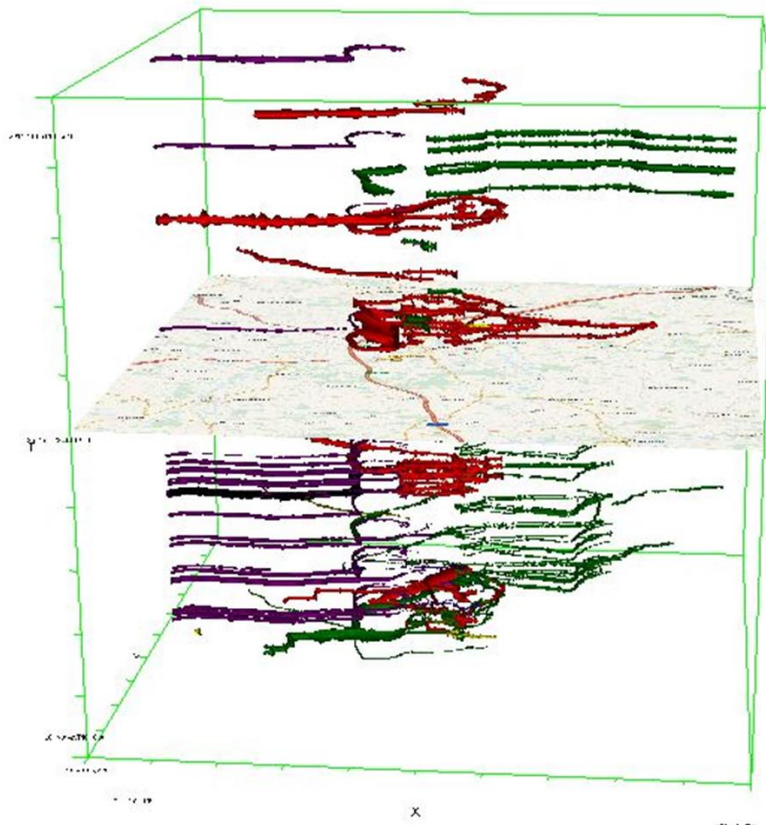
Single track zoomed



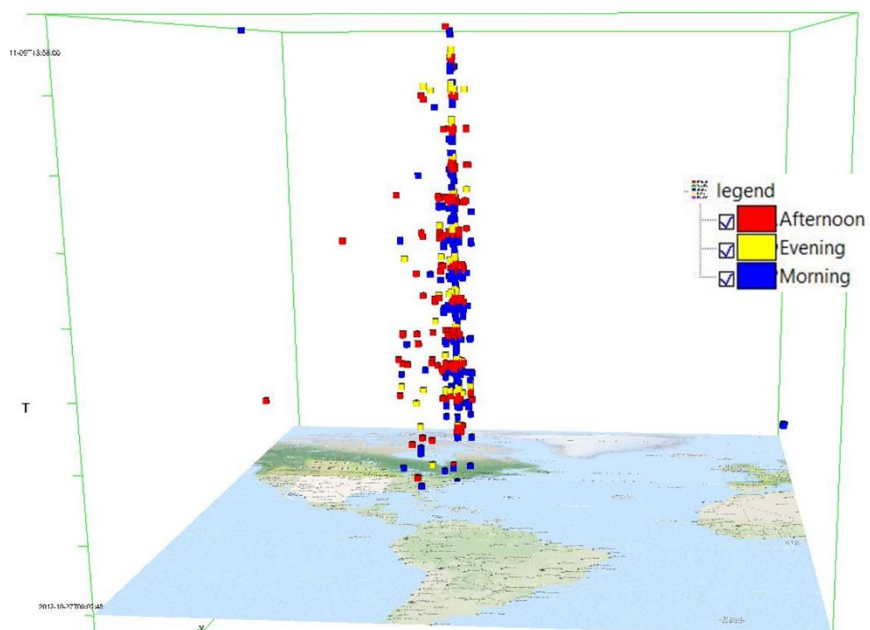
Top View of STC



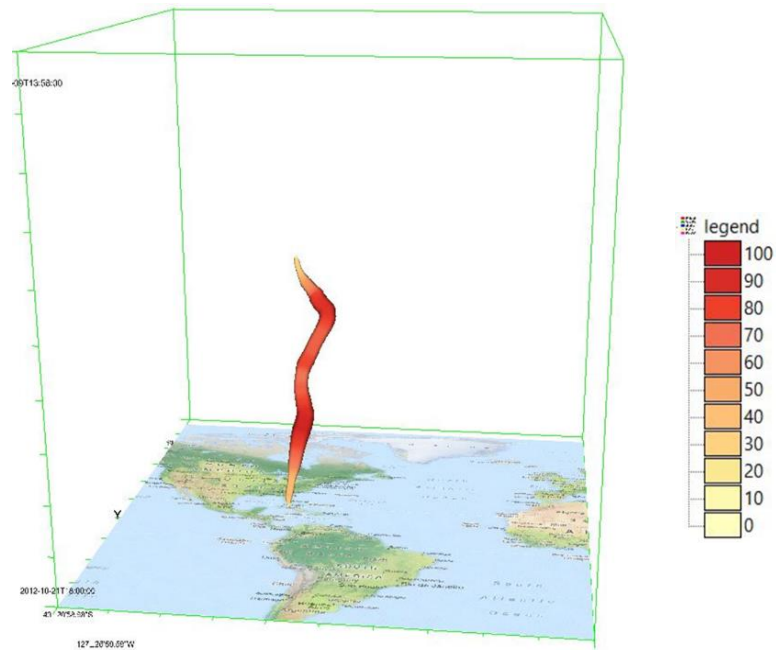
Side View of STC



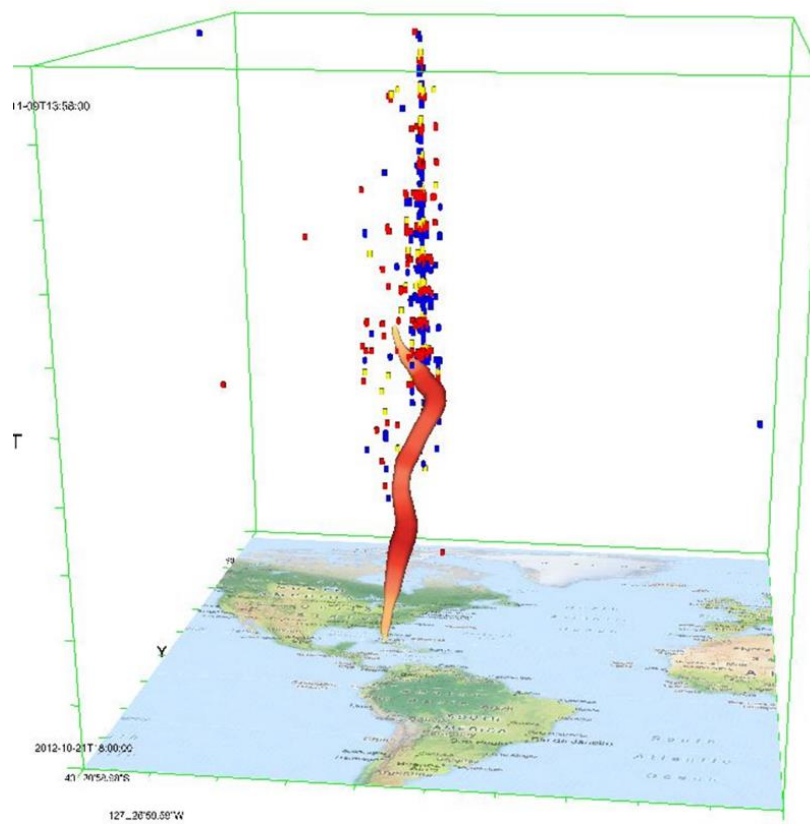
Change in base map for STC



Flickr posts



Sandy Hurricane Track in STC



Flickr posts along with Sandy Hurricane track

APPENDIX 3: CODE USED TO DOWNLOAD ENVIROCAR DATA

```
# encoding: utf8

"""
Download script for envirocar measurements
for a specific bounding box.
It loads all measures into RAM before
writing the output. So be warned before
making the bounding box too large.
"""

import requests
import json
import csv

BBOX = (7.379028,51.823596,7.903453,52.087036)

#LIMIT = 3
LIMIT = 10000

def get_measurements(page):
    """
    Downloads one page worth of measurements
    """
    url = "https://envirocar.org/api/stable/measurements"
    params = {
        "bbox": "%s,%s,%s,%s" % BBOX,
        "page": page,
        "limit": 100
    }
    r = requests.get(url, params=params)
    for feature in r.json()["features"]:
        yield feature

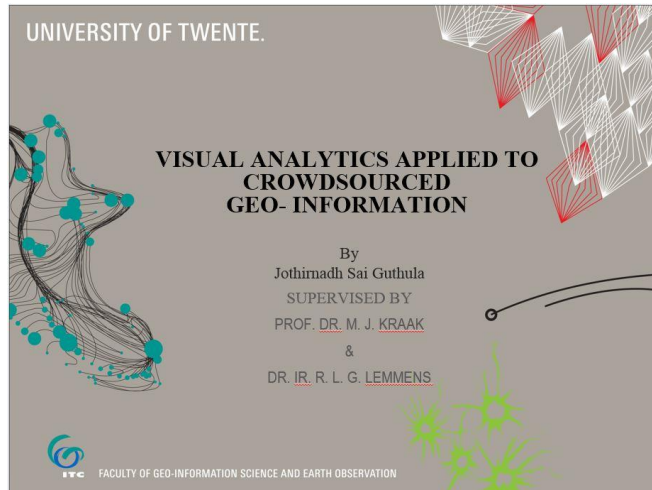
def get_phenomenons():
    url = "https://envirocar.org/api/stable/phenomenons"
    r = requests.get(url)
    return r.json()["phenomenons"]

if __name__ == "__main__":
    phenomenons = sorted(get_phenomenons(), key=lambda k: k["name"])
    measures = []
    for page in range(1, LIMIT):
        mlist = list(get_measurements(page))
        if len(mlist) == 0:
```

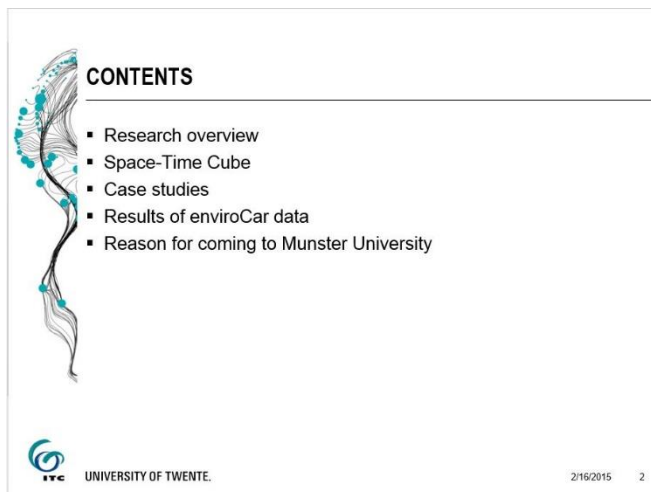
```
        break
    for m in mlist:
        measures.append(m)
measures = sorted(measures, key=lambda k: k["properties"]["id"])
with open("measures.csv", "wb") as csvfile:
    writer = csv.writer(csvfile)
    # header row
    headers = [
        "id",
        "datetime",
        "lon",
        "lat",
        "sensor_type",
        "sensor_id",
        "sensor_manufacturer",
        "sensor_model",
        "sensor_fueltype",
        "sensor_construction_year",
        "sensor_engine_displacement"
    ]
    for p in phenomenons:
        headers.append(p["name"])
    writer.writerow(headers)
    # data rows
    for m in measures:
        row = [
            m["properties"]["id"],
            m["properties"]["time"],
            "%.6f" % m["geometry"]["coordinates"][0],
            "%.6f" % m["geometry"]["coordinates"][1],
            m["properties"]["sensor"]["type"],
            m["properties"]["sensor"]["properties"]["id"],
            m["properties"]["sensor"]["properties"]["manufacturer"],
            m["properties"]["sensor"]["properties"]["model"],
            m["properties"]["sensor"]["properties"]["fuelType"],
            str(m["properties"]["sensor"]["properties"]["constructionYear"]),
            ### m["properties"]["sensor"]["properties"]["engineDisplacement"],
        ]
        for p in phenomenons:
            if p["name"] in m["properties"]["phenomenons"]:
                row.append(str(m["properties"]["phenomenons"][p["name"]]["value"]))
            else:
                row.append("")
        writer.writerow(row)
```

APPENDIX 4: PRESENTATION SLIDES

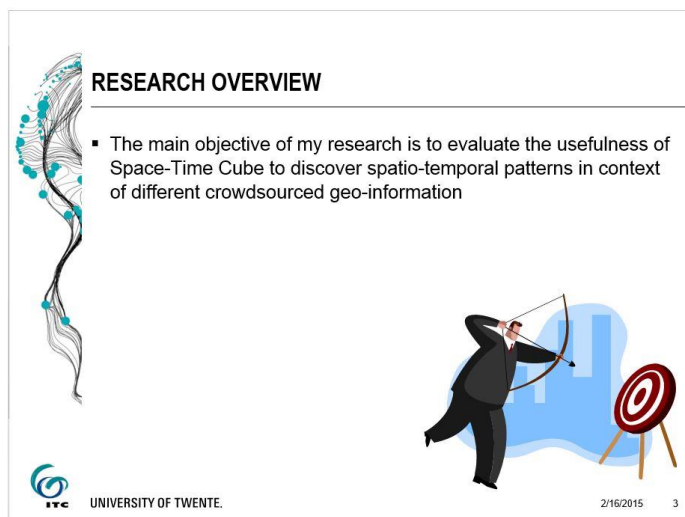
1.



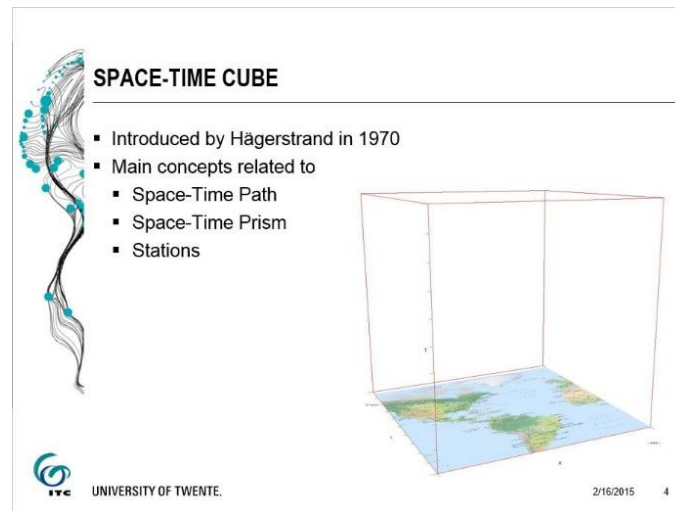
2.



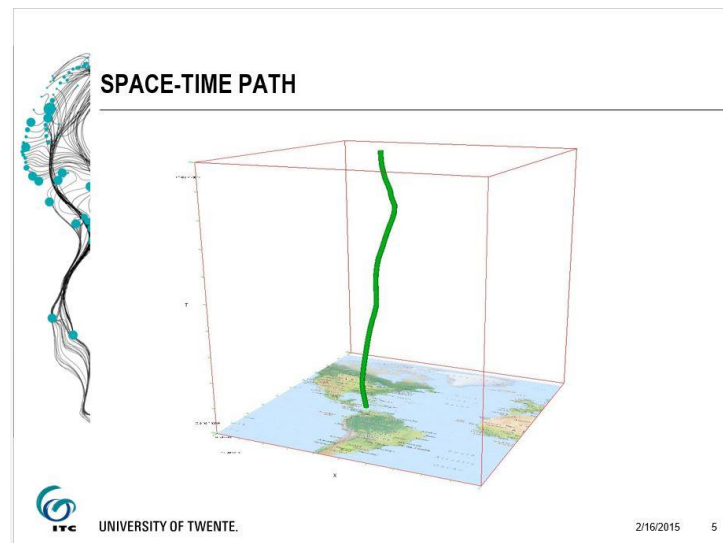
3.



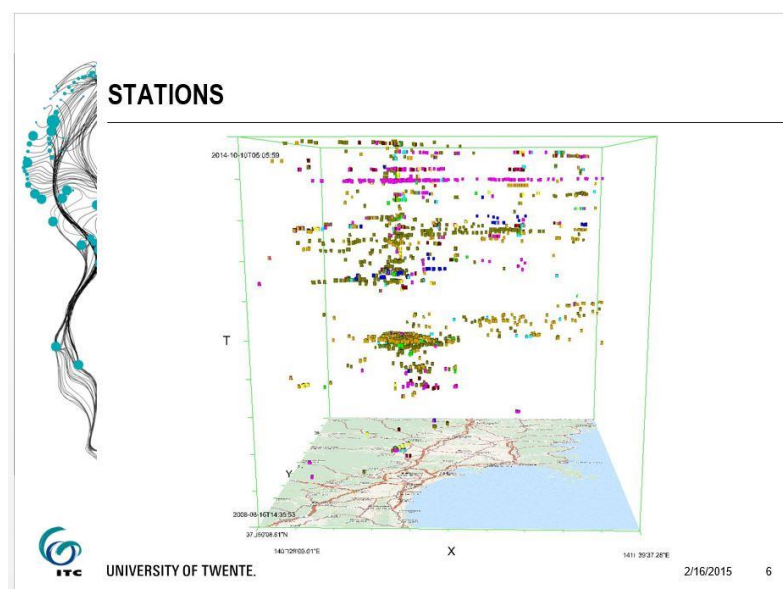
4.



5.



6.



7.



CASE STUDIES


Three case studies

- Flickr
- OpenStreetMap
- enviroCar

ITC UNIVERSITY OF TWENTE.

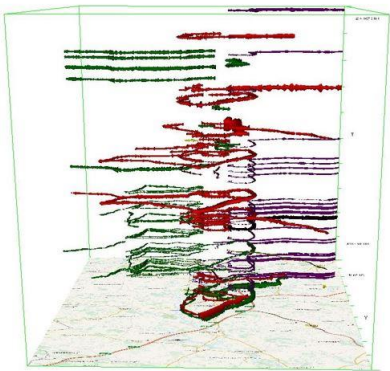
2/16/2015 7

8.



RESULTS OF ENVIROCAR DATA


- Demonstrate this in ILWIS software



ITC UNIVERSITY OF TWENTE.

15 8

9.



REASON FOR COMING TO MUNSTER UNIVERSITY

- Feedback and evaluation
 - Usability analysis form

ITC UNIVERSITY OF TWENTE.

2/16/2015 9

LIST OF REFERENCES

- Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., ... Tominski, C. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10), 1577–1600. doi:10.1080/13658816.2010.508043
- Arsanjani, J. J., & Bakillah, M. (2014). Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap. *International Journal of Digital Earth*, 1–16. doi:10.1080/17538947.2014.951081
- Bertin, J. (1983). *Semiology of graphics*. (W. J. Berg, Ed.). The University of Wisconsin Press. Retrieved from <http://catalogue.nla.gov.au/Record/2340325>
- Demšar, U., & Verrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, 24(10), 1527–1542. doi:10.1080/13658816.2010.511223
- Díaz, L., Granell, C., Huerta, J., & Gould, M. (2012). Web 2.0 Broker: A standards-based service for spatio-temporal search of crowd-sourced information. *Applied Geography*, 35(1-2), 448–459. doi:10.1016/j.apgeog.2012.09.008
- Dickinson, J., Shirk, J., & Bonter, D. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology* Retrieved from <http://www.esajournals.org/doi/abs/10.1890/110236>
- Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, 102(3), 571–590. doi:10.1080/00045608.2011.595657
- enviroCar. (2013). Retrieved May 23, 2014, from <https://envirocar.org/mapsandstatistics.php>
- Goodchild, M. F. (2007a). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221. doi:10.1007/s10708-007-9111-y
- Goodchild, M. F. (2007b). Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research*, 2, 24–32. Retrieved from <http://www.geoinformatics.cn/wp-content/uploads/citizensasvoluntarysensors.pdf>
- Goodchild, M. F. (2012). VGI Research Overview. Retrieved from http://web.ornl.gov/sci/gist/workshops/2012/vgi_documents/2012_VGIworkshop_Goodchild.pdf
- Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24, 6–21.
- Haklay. (2007). Muki Haklay's personal blog. Retrieved December 11, 2014, from <http://povesham.wordpress.com/>
- Haklay, N. B. (2010). OpenStreetMap—Overview and Motivational Factors.
- Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 550–557. doi:10.1016/j.isprsjprs.2010.06.005

- Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business 1*. Crown Publishing Group New York, NY, USA ©2008. Retrieved from <http://dl.acm.org/citation.cfm?id=1481457>
- Hudson-Smith, a., Batty, M., Crooks, A., & Milton, R. (2009). Mapping for the Masses: Accessing Web 2.0 Through Crowdsourcing. *Social Science Computer Review*, 27(4), 524–538. doi:10.1177/0894439309332299
- Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (Eds.). (2010). *Mastering the information age : solving problems with visual analytics* (p. 168). Visual Analytics - Mastering the Information Age (VisMaster). Retrieved from <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf>
- Kraak, M. J. (2003). The space - time cube revisited from a geovisualization perspective. In *International cartographic conference* (pp. 1988–1996). ITC. Retrieved from http://www.itc.nl/library/Papers_2003/art_proc/kraak.pdf
- Kraak, M. J., & Madzudzo, P. (2007). Space Time Visualization for Epidemiological Research, 9.
- Kraak, M. J., & Ormeling, F. (2011). *Cartography: Visualization of Spatial Data* (Third edit., p. 198). Harlow, Pearson Education Ltd.
- Kveladze, I., Kraak, M.-J., & van Elzakker, C. P. J. M. (2013). A Methodological Framework for Researching the Usability of the Space-Time Cube. *The Cartographic Journal*, 50(3), 201–210. doi:10.1179/1743277413Y.0000000061
- Kwan, M.-P. (2000). Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8(1-6), 185–203. doi:10.1016/S0968-090X(00)00017-6
- Ludicorp. (2004). Flickr. Retrieved November 11, 2014, from <https://www.flickr.com/about>
- MacEachren, A. M., Robinson, A. C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J., & Mitra, P. (2011). Geo-Twitter Analytics: Applications in Crisis Management (p. 8).
- MacEachren, A. M. (1994). *Some Truth With Maps: A Primer on Symbolization & Design*. Resource Publications in Geography.
- Miller, H. J. (1999). Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures. *Geographical Analysis*, 31(2), 187–212. Retrieved from <http://doi.wiley.com/10.1111/j.1538-4632.1999.tb00976.x>
- Monmonier, M. (1993). *Mapping It Out: Expository Cartography for the Humanities and Social Sciences*, Monmonier. The University of Chicago Press. Retrieved from <http://www.press.uchicago.edu/ucp/books/book/chicago/M/bo3645221.html>
- Neis, P., & Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(3). doi:10.3390/ijgi1020146
- Nielsen, J. (1993). *Usability Engineering*. Academic Press, Boston.
- Nielsen, J. (2006). Participation Inequality: The 90-9-1 Rule for Social Features. Retrieved December 12, 2014, from <http://www.nngroup.com/articles/participation-inequality/>

- OpenScientist. (2011). Finalizing a Definition of “Citizen Science” and “Citizen Scientists.” Retrieved December 11, 2014, from <http://www.openscientist.org/2011/09/finalizing-definition-of-citizen.html>
- OpenStreetMap. (2004). Retrieved August 15, 2014, from <http://www.openstreetmap.org/#map=12/38.9020/-77.0248>
- Peuquet, D. J. (1994). It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers*, 84(3), 441–461. doi:10.1111/j.1467-8306.1994.tb01869.x
- Ramm, F. (2009). Crowdsourcing Geodata. In *Looking back on 5 years of OpenStreetMap* (p. 33). Proceedings FOSSGIS 2008, Freiburg, <http://www.geofabrik.de/media/2009-09-08-crowdsourcing-geodata.pdf>. Retrieved from <http://www.geofabrik.de/media/2009-09-08-crowdsourcing-geodata.pdf>
- Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of “big data.” *Geoforum*, 52, 167–179. doi:10.1016/j.geoforum.2014.01.006
- Surowiecki, J. (2004). *The Wisdom of Crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York: Little, Brown and Company. Retrieved from <http://www.amazon.com/The-Wisdom-Crowds-James-Surowiecki/dp/0385721706>
- Thomas, J. J., & Cook, A. K. (Eds.). (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics* (p. 184).
- Tsou, M.-H., & Leitner, M. (2013). Visualization of social media: seeing a mirage or a message? *Cartography and Geographic Information Science*, 40(2), 55–60. doi:10.1080/15230406.2013.776754
- Wasserman, S., & Faust, K. (1995). *Social Network Analysis: Methods and Applications* (p. 857). Cambridge University Press. Retrieved from <http://www.cambridge.org/nl/academic/subjects/sociology/sociology-general-interest/social-network-analysis-methods-and-applications>
- Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy*, 2(2), 6–32. doi:10.2202/1948-4682.1069

URL

1. <https://www.youtube.com/watch?v=ZmqR-8sFAEQ>
2. <https://www.youtube.com/watch?v=kFwdtp52MS4>
3. <https://www.youtube.com/watch?v=V7-7zVyRXIY>
4. <https://www.flickr.com/services/api/explore/flickr.photos.search>
5. <http://usability.ru/sources/iso9241-11.htm>