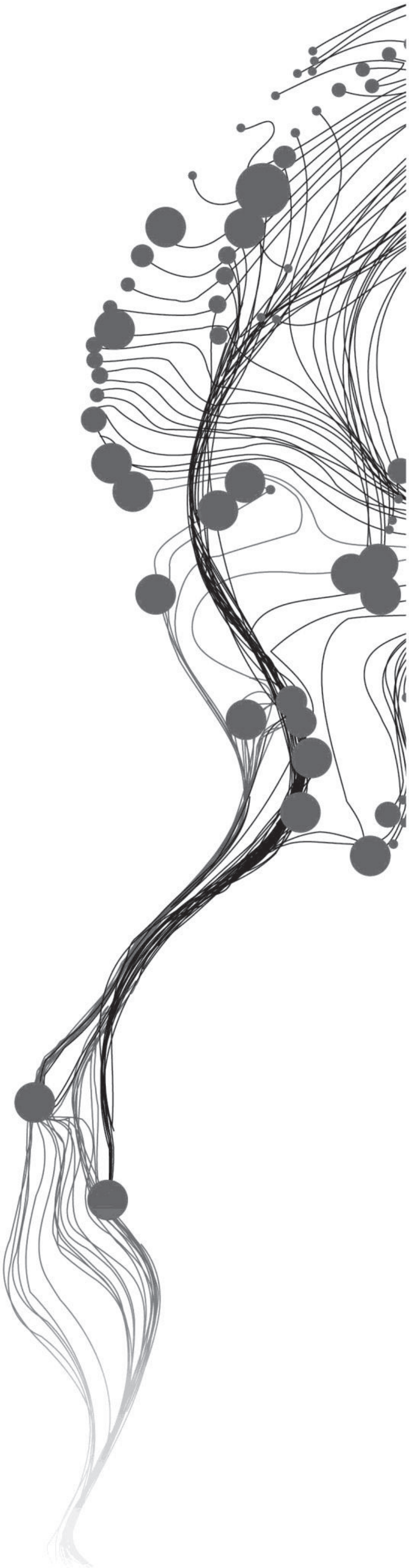
COMPARISON OF VISUAL OUTPUTS OF TWO SPATIOTEMPORAL CLUSTERING TECHNIQUES AND THE RELEVANT USER RESEARCH

QIAN LIU
February, 2014

SUPERVISORS:
Prof.Dr. M.J. Kraak
Dr. C.P.J.M. van Elzakker

Advisor:
Xiaojing Wu MSc
Dr. R. Zurita-Milla



COMPARISON OF VISUAL OUTPUTS OF TWO SPATIOTEMPORAL CLUSTERING TECHNIQUES AND THE RELEVANT USER RESEARCH

QIAN LIU
Enschede, The Netherlands, February, 2014

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: [Name course (e.g. Applied Earth Sciences)]

SUPERVISORS:
Prof.Dr. M.J. Kraak
Dr. C.P.J.M. van Elzakker

Advisor:
Xiaojing Wu MSc
Dr. R. Zurita-Milla

THESIS ASSESSMENT BOARD:
Dr. A.A. Voinov (Chair)
Dr. Ir. R.J.A. van Lammeren (External Examiner , Wageningen UR)

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

Spatiotemporal big data processing and analyzing is a popular topic in recent years. Spatiotemporal big data contains a large amount of information about time, location and attributes; making use of this information can help people better understand the world. Hence, a suitable technique is the key to be applied to this kind of spatiotemporal big data to obtain information. Some of the clustering techniques make it possible to discover information from spatiotemporal big data; different clustering techniques reflect different types of information. Thus, choosing a suitable clustering technique for a given spatiotemporal dataset is very important, and this leads to comparisons between different clustering techniques.

This research describes the suitability of two clustering techniques (SOM and co-clustering) to answer general questions related to spatiotemporal big data by means of their visual outputs, based on a case study spatiotemporal big data, which is the Netherlands' 20-year daily temperature data from 28 Dutch meteorology stations. The generic questions are generated based on the nature and characteristics of this kind of big data, and then they are organized and summarized; the typical visual outputs from the two selected clustering techniques and case study dataset were selected and analyzed for answering generic questions, and the analysis result are organized, summarized and formulated into hypotheses (about what kind of questions can be answered by which visual outputs). For the purpose of testing the hypotheses, domain questions that reflect the real information requirements are requested from domain experts, based on the case study dataset. They are organized and summarized in the same way as generic questions, and are verified as consistent with the generic questions so that the answers to the domain questions can used to check the answers to the generic questions. An experiment which involved nine test participants and several tasks was designed and executed to define the strengths and weaknesses of the selected clustering visual outputs. The tasks for test participants included a question-solving task (participants answer domain questions by using selected visual outputs), a think aloud task and a questionnaire. The user research methods used by the investigator to evaluate the results of the tasks are: observe, record (includes taking notes, video recording and time registering) and focus group. By using the answers to the domain questions, the hypotheses are tested; by analyzing the correctness, time recording and questionnaire, the effectiveness, efficiency and satisfaction are described.

Key words: Spatiotemporal big data, SOM, Co-clustering, Visual outputs, Hypotheses, User test

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank all the people who supported me during my MSc study at ITC, particularly Lichun Wang, Tina Tian and those in the department of Geo-information Processing.

First of all, my sincere gratitude goes to my two supervisors, Prof. Dr M.J. Kraak and Dr. C.P.J.M. van Elzakker , who have been of great help and guidance. This thesis could not have been written without their support and contributions. My sincere thank also goes two my two advisors, for the discussions and guidance.

I would like to thank all my test participants, without whose understanding and support, I could not have finished the research.

My grateful thanks to all my lovely classmates and friends at ITC, especially to Eden Mashilla, Immaculate Katutsi, Gewa Li, Guorui Li, Mafkereseb Kassahun, Sihhdi and Rihhdi Munde, Gustavo Garcia, Manuel Garcia and my six group members for their encouragements, friendship and support; without them my study and research life in ITC would not have been as colourful.

My deep thanks also go to my good friends outside of ITC, John Esposito, Alastair Burt, Shiqiao Liu, Xin Liu, David Qin, Tom, and Ivan Maggiori, thanks for all the encouragement and support.

I cannot finish this acknowledgement without mentioning my parents, who have been accompanying me with their endless love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Motivation and problem statement

The explosion of data gathered from a variety of sources nowadays results in the big data problem. It is difficult to derive useful information or understand the big data directly due to its large data size, high dimension and information overloading (Fu, 2011). Understand big data might help better understanding the world around us. The first step towards understanding these big data is to identify trends, characteristics and correlations among the data objects. The findings can then be visually presented to the user. Especially, the analysis of spatiotemporal data is much more challenging than other common big data sets because it includes both geospatial and temporal information (Andrienko et al., 2010). It is widely known that using either an automatic machine method or human ability for visual analysis, it is hard to achieve the purpose of the application.(Andrienko & Andrienko, 2013). A combination is required, and finding a suitable method to represent and analyze spatiotemporal data has become an important research task.

In this situation, the clustering technique is regarded as a fundamental data processing method to discover information (Deng et al., 2011; Jain et al., 1999). Because it integrates data categorization and visualization, the clustering technique is widely used in data mining, pattern recognition, image analysis, business surveys, etc.(Lu & Wan, 2012; Rokach & Maimon, 2005). The main idea of the clustering technique is to group similar objects together, while different objects are clustered into different classes (i.e. to make the maximum similarity of objects in the same group, and the maximum difference between objects in different groups). Consequently, data is grouped into different groups and the results are represented in different graphic visual outputs. The visual outputs of the clustering technique are key in observing, analyzing and extracting information from big data. It enables us to discover a new set of categories (Rokach & Maimon, 2005), giving us an insight into big data sets by presenting data graphically.

There is a wide range of clustering methods, but a deficiency of information to describe how each clustering method works and what kind of problems it can solve. One of the main reasons of the diversity of clustering methods is because there is no precise definition of the notion of "cluster". Previous researchers have proposed many algorisms; these algorisms are "mathematical formalizations of what researchers believe is the definition of cluster" (Estivill-Castro, 2002). The researchers' backgrounds are different, as are the applications. These factors bring about biased clustering algorithms, resulting in variations in the clusters. Thus, for the same reason, each clustering algorithm has been applied to different types of datasets for different purposes. Therefore, it is impossible to determine the "optimal" one. The data processing results of clustering methods are displayed in their visual outputs, representing different aspects of the information of the dataset. For one special purpose, it is possible to determine which method is capable of representing which aspect of information from a given data set according to the experimental studying of clustering visual outputs. Thus, with the purpose in mind, selecting a clustering method based on the visual outputs for representing information from a specified dataset at hand is an important task in the data analysis field.

Based on the above discussion, this research is motivated towards helping researchers to choose the clustering method based on its visual outputs for a particular purpose; and offering detailed information about the performances of the chosen clustering visual outputs in this research, whose possible deriving principles that could be applied more generally to other data sets.

1.2. Research identification

A clustering method for a particular data set is commonly chosen experimentally and subjectively because "clusters are, in large part, in the eye of the beholder" (Estivill-Castro, 2002), and it is hard to use mathematical evidence to show which clustering method is suitable for an application. This has prompted researchers to focus on the comparison of the various clustering methods from different perspectives. Earlier research compared the clustering methods from the algorithm view (Hirano & Tsumoto, 2005; Lai et al., 2012; Meila & Heckerman, 2001).

In this research we concentrate on the comparison of visual outputs of selected clustering techniques base on a case study dataset, to offer an alternative approach for selecting proper clustering methods for exploring big spatiotemporal dataset for potential users. This approach is supposed to be an intuitive way compared to other approaches.

Based on the literature review, we have selected two clustering methods (i.e. SOM and Co-clustering), because they are efficient and typical for being used in big spatiotemporal datasets.

The SOM method is often used for spatial time series data clustering (Hudson et al., 2011; Wang et al., 2006). Co-clustering is a prevalent technique that has good performance in managing high-dimensional data (Wu et al., 2012).

1.3. Research objectives

The main objective of this research project is to describe the suitability of two clustering methods (i.e. SOM and Co-clustering) to answer questions related to spatiotemporal big data, by means of their visual outputs, based on the case study spatiotemporal weather dataset.

This main objective can be subdivided into the following sub-objectives

A. To define generic spatiotemporal questions to address big datasets.
B. To establish the characteristics of the visual outputs that generated from the two selected context dependent clustering techniques.
C. To find out what type of questions experts have in a particular context working with big data, such as spatiotemporal temperature data.
D. To evaluate the strength and weakness of the visual output of the chosen clustering methods in answering the experts' questions.

1.4. Research questions

Related to sub-objective A

1. What are the generic questions of spatiotemporal big dataset?
2. How to organize the generic questions so that can summarize the answers to these questions concisely?
3. What are the spatial and temporal characteristics of the case study dataset?

Related to sub-objective B

4. What visual outputs are generated by SOM and Co-clustering on the basis of the case study dataset are selected?
5. How do the selected visual outputs answer the generic questions?
6. Based on the selected case study visual outputs, what hypotheses can be generated in relation to generic spatiotemporal questions?

Related to sub-objective C

7. What kind of questions do domain experts have with the case study dataset?
8. How can the domain questions be organized similar to the generic questions and can be used for the test?

Related to sub-objective 3

9. How to set up and execute experiments to define the strengths and weaknesses of visual outputs in answering experts' questions?
10. How do users answer the specific domain questions with the help of the visual outputs of the clustering techniques?
11. How to compare test results with hypothesis to draw conclusion regarding the strength and weakness of visual outputs?

1.5. Thesis structure

This thesis includes seven chapters. Chapter two is devoted to the reviews of the literature about background, outlook and characteristics of big data, specifically the type we focused on this research study, spatiotemporal data; the overview of clustering technique, the theoretic descriptions of how the selected clustering methods work and their common used visual representations. Chapter three describes the methodology applied during this research study. Chapter four describes the case study, selects and presents the visual outputs used. Chapter five is about visual analytics which is based on the generic questions and the selected visual outputs, and formulate the hypotheses on the basis of the visual analysis result; in the end, collect the domain questions for checking the generic questions and preparing the hypotheses test. Chapter six is devoted to a user evaluation presenting the experiment set up, execution and the experimental result. Chapter seven contains integration analysis of the whole research, final conclusion, along with recommendations to the potential users for further research.

2. EXPLORING BIG SPATIOTEMPORAL DATASETS THROUGH CLUSTERING

2.1. Introduction

This chapter presents the state of the art regarding the exploring big data, especially spatiotemporal data through clustering techniques. It starts with introducing the background and the outlook of big data, discuss the meaning of the big data research, then put the attention on the special type- big spatiotemporal data, which we are focus on in this research. And then elicit the clustering technique that used for exploring the big data. Finally, the two clustering methods are selected that are supposed to be able to handle the spatiotemporal dataset as focuses and their common used visual representations are described.

2.2. Big data

We are now living in a big data driven world. All the web services, healthcare, public sector, retail, manufacturing and personal location services cannot operate effectively without big data (Sagiroglu & Sinanc, 2013). Big data is not a precisely defined term; it usually refers to the data sets with complex structure and large data size that overwhelm the commonly used software or human ability alone. It has three main components: variety, velocity and volume, which refer to data type, data update speed and data size, respectively. The data is exploding rapidly and largely. It is calculated that 2.5 quintillion bytes of data is created per day and 90% of the data in the world today are from only the last two years; and it is considered to be an opportunity for finding insights in new emerging data types to answer questions that were previously considered but out of reach (Dobre & Xhafa). The big data is usually difficult to store, analyze and visualize for further use. Spatiotemporal datasets, especially time series spatiotemporal data sets as a special type of big data, receiving much more attention in the data mining field in recent years (Wang et al., 2006).

2.2.1. Spatiotemporal data

Spatiotemporal data deals with spatial, temporal and attributes of complex objects. With the time dimension, people can find the changes of the spatial data and understand the dynamic relationships in the data (Wei & Hou-kuan, 2006). The analysis and visualization of spatiotemporal data usually include the identification of patterns, trends and correlations of the data elements over time, also taking into account the heterogeneity of the space and structure of time; the visualization often plays a key role in the successful analysis of time-related data (Keim et al., 2008).

Again, because of the characteristics of large volume, high dimension, information overload and continuous updating of this kind of dataset, we need to find a strategy to retrieve the information that we need from it. For this, we need to combine geo-computational software and human cognitive ability.

The different data mining tasks relevant to represent the time series are mainly in four fields: classification, rule discovery, summarization, pattern discovery and clustering (Fu, 2011). The focuses of this research are the pattern discovery and clustering.

2.3. Techniques of clustering

No matter what kind of big data needs to be analyzed, data categorization, aggregation and representation are necessary steps for finding meaningful structure to analyze. The clustering technique is an exploratory data analysis tool that could be used to discover structure in big data. As more and more big data need to be analyzed, clustering exists in almost every aspect in our daily life.

Over time, more than one hundred clustering methods have been developed. Kleinberg (2003) proposing his impossibility theorem, indicating that there is not a single clustering algorithm that in all circumstances performs better than the rest of the algorithms. This explains the wide range of existing clustering algorithms. The diversity of clustering methods contributes to the diversity of clustering results and clustering visual outputs , and "any representation will inevitably favor an interpretation over all possible ones" (Keim et al., 2008). The fact that researchers have developed different clustering algorithms made several researchers compare the clustering methods from multiple perspectives. Algorithms were compared in (Meil, 2005) and (Hossain et al., 2012; Kleinberg, 2003), and also the performance criteria (model structure, cluster quality and running time etc.) were compared from the algorithm perspective. Some good surveys and comparisons in the literature even looked the comparison methods and reported the characteristics of clustering methods based on the comparison (Hirano & Tsumoto, 2005), or compared the effectiveness of clustering algorithms (Marinai et al., 2008). So far, none of the reviewed literature compared clustering methods from a visual output perspective, and none of them considered a comparison of the clustering visual outputs in answering specific questions, and they also didn't compare the two clustering methods that we selected in this research project. The great varieties of clustering methods are standard tools in many data mining system. But for a complex data type, for instance spatiotemporal data, it requires special methods that can group time series with similar structures together and address one or more of the below problems: data size reduction, data dimension reduction, trend and characteristic discovery (Bock, 2011).

2.3.1. SOM

The term self-organizing maps (SOM) or the Kohonen map was first proposed and explained in detail by Teuvo Kohonen in his book Self-Organizing Maps in 1995. The self-organizing map is a type of artificial neural network, based on the neural network theory, artificial and biological. Inside the SOM, the input data is trained unsupervised; the output nodes are formed on the basis of the similarity of the input data. Via this unsupervised learning procedure, similar input data will be in the same output node, and dissimilar ones will be separated. In this way, SOM always performs dimension reduction while clustering, which simplifies the complexity; in the meantime the outputs nodes reveal the meaningful structures and relationships that exist but are invisible in the input space (Mount, 2009). As a clustering technique, SOM has been used for processing high dimensional time series in data mining tasks, and usually displays high-dimensional data in simple two-dimensional space (Wang et al., 2008; Wang et al., 2006). Many SOM research studies have been done on data clustering (Agarwal & Skuping, 2008; Wu & Chow, 2003).

Because SOM is data driven, the output and data structure can be different depending on the input data and it keeps the topology relationship of mapping from the input to output space (Mount, 2009). For exhibiting the SOM clustering result, there are several forms of representations that are commonly used.
The U-matrix map (the unified distance matrix) is a representation that is used in SOM to depict the distance between neighbour neurons of the network visually. The distances are calculated from the centre of one neuron to all of its neighbours by using different values of colouring. A dark colouring shows the large distance between neighbour neurons and that represents the value differences in input space, while the light colouring implies short distances between neighbour neurons, which means the values are close to each other in the input space, and the neurons that have similar values are close to each other. This

representation is easy for researchers to separate the whole map into several clusters. The u-matrix map can uncover and give us an impression of the data structure in datasets (Koua, 2005).
The Cluster map is generated based on the U-matrix map. It groups similar nodes (according to colour) on the U-matrix map inside the same cluster. Thus, from the nodes distribution, we can see the relationship between input objects.

The Geographical map is used to project SOM results, so that we can see the spatial distribution of the clusters. This helps us find and understand the data patterns in a geographic view, and assists in detecting abnormal patterns that are relevant to the geographic location and explain the phenomena according to the geographic reality.

The Trend plot is a plot that uses lines to represent the trend of objects in a dataset (the objects are clustered into different clusters). Trend plots assist in visualizing the SOM results, getting the insights from the dataset. This visualization can be used to detect the trajectories of phenomena over time.
The Anomalies graph shows the obvious differences between objects in the same cluster. This is because the cluster's value is different from the values of other objects in the same cluster.

2.3.2. Co-clustering

Given a data matrix table, with rows and columns (rows and columns are different aspect of objects). SOM and other clustering methods can help retrieve different aspects of data information by either clustering the rows of the data table or clustering the columns, successively and independently. But if we are interested in knowing about both the row and the column information at the same time, co-clustering can deal with this issue. The co-clustering, also called bi-clustering or two-mode clustering (Van Mechelen et al., 2004), is another important unsupervised clustering technique, first introduced by Mirkin (1996). It is a two-way clustering algorithm for generating co-occurrence statistics, presented as a novel method which simultaneously clusters data from both rows and columns of a contingency table simultaneously (Dhillon et al., 2003). In this paper, the authors used a word document as an example, demonstrating that the co-clustering method "worked well in practice, especially in the presence of high-dimensionality". Moreover, the co-clustering method was used by Moghaddam et al. (2010) for 79 locations, thousands of mobile users, and billions of records of spatiotemporal data; the result showed that the co-clustering method is efficient at both the global and location-based levels.

The density map (also called heat map) and geographical map are often used as co-clustering visual outputs. A heat map is basically a table with colours instead of numbers. It is a direct way of visualizing the values contained in a data matrix. The colour scale depends on the measurement level of values. Rows and columns are ordered, so that similar rows and columns are next to each other, displaying the cluster groups (Wilkinson & Friendly, 2009). So the intersection of rows and columns show the co-occurrence information. The geographical map used in a co-clustering representation is similar to that used in the SOM. It represents the information in a heat map in a spatial environment.

2.4. Conclusion

Based on the reviews, the relevant descriptions of the two selected clustering theories and research studies show that the two clustering methods have the ability to handle spatiotemporal datasets.
The visual representations of the two clustering methods reveal different aspects of the hidden information in the dataset. Aiming to understand the information the different clustering visual outputs offering, we need to use some strategy to explore the visual outputs.

3. METHODOLOGY OF RESEARCH STUDY

3.1. Introduction

This chapter represents the research workflow, describes the framework of methodology and the theoretical approach, along with the usability methods.

3.2. Framework of methodology

This research is dedicated to finding out what kinds of questions, related to the spatiotemporal dataset, can be answered by the visual outputs of the selected two clustering techniques. The result can be recommended to potential users, who may choose a clustering technique to process spatiotemporal data. Meanwhile, we want to describe the performance of the selected visual outputs.

A set of generic questions are generated from studying the theory of spatiotemporal big data, and bring in a spatiotemporal case study big dataset. On the one hand, based on the case study data and the selected clustering techniques, we select clustering visual outputs; on the other hand, the real information requirements (specific domain questions) are generated from the case study dataset, by asking domain experts to list their possible questions according to the case study spatiotemporal dataset.

Generic questions are used to analyze the selected clustering visual outputs, examining what kind of generic questions the selected clustering visual outputs can answer; according to the result of the visual analysis, hypotheses are formulated about which clustering visual outputs are suitable for answering what kinds of questions. Then experiments are set up, to test the hypotheses and get the in-depth performance information of the clustering visual outputs. Specific domain questions are used in the experiment for testing hypotheses, by finding the connection between the generic questions and the domain questions, and then using test users to find the answers to the domain questions in the selected clustering visual outputs to see if the answers to the two sets of questions are consistent. In addition to this information-seeking task, several other user research methods are used in the experiment in order to gain in-depth information about the performance of the selected clustering visual outputs.

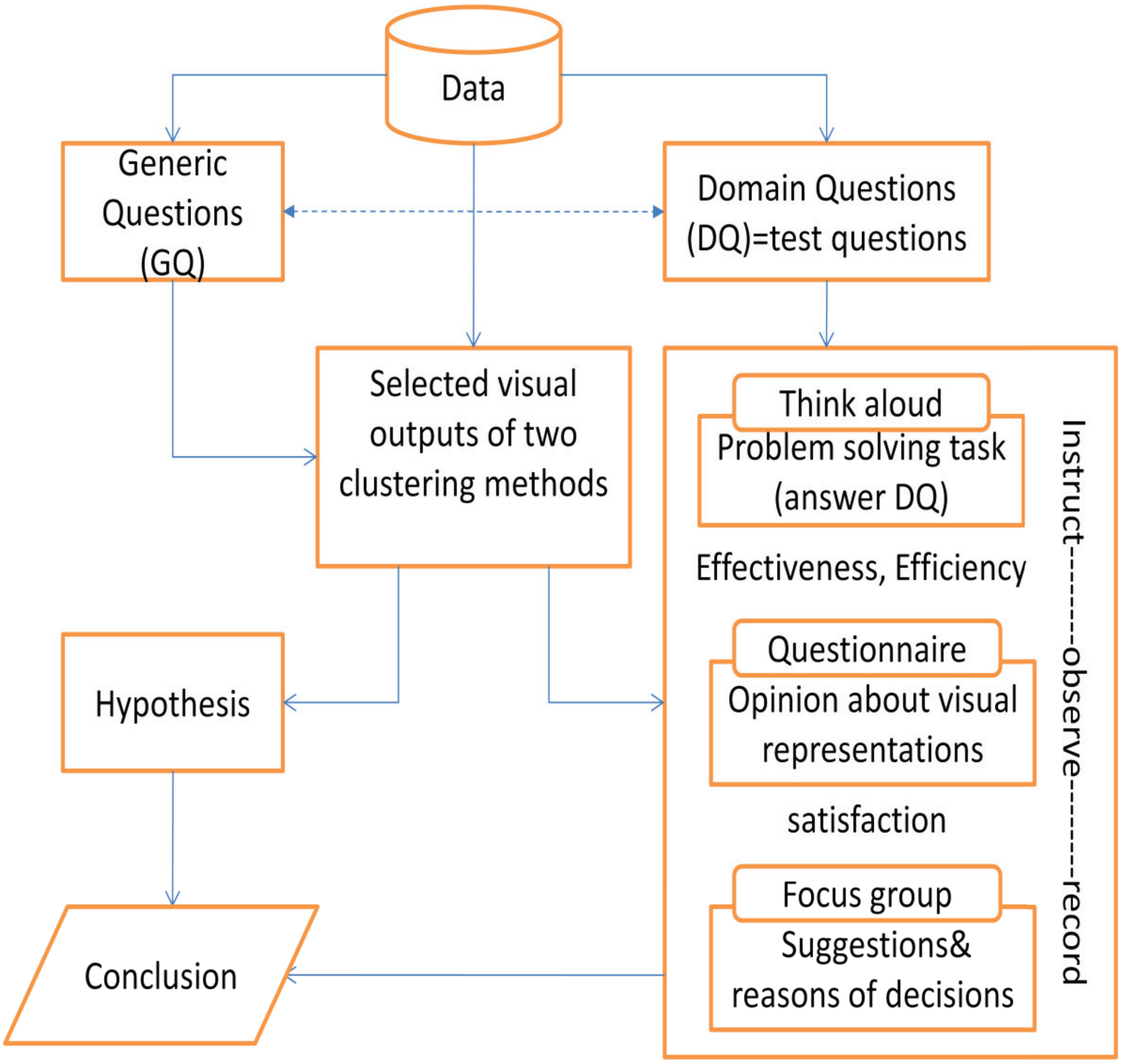


Figure 3.1Work flow

3.3. Approach

This section describes the approaches that are mentioned in the framework.

3.3.1. Generate generic questions and select case study data

To achieve the first sub-objective, the first step is to generate generic questions and then select a case study dataset. The generic questions are generated based on the nature and inherent properties of the spatiotemporal dataset. Because the analysis of spatiotemporal data usually involves three components: where (location), when (time), and what (attribute/thematic objects) (Peuquet, 1994), we propose the generic questions from these three components. Also the questions are supposed to cover different information search levels, which are elementary, intermediate and overall (Andrienko et al., 2003; Bertin, 1983), from local to global.

Research questions 1 and 3 will be answered in this phase.

3.3.2. Organize generic questions

From the observation of Andrienko et al. (2003), we learned that the exploration of an intermediate search level usually doesn't differ much from exploring an overall search level, so this paper categorizes search levels into only two categories, general and elementary. Following this approach, when the generic questions are generated, all the questions will be analyzed and then the similar questions that differed in the intermediate and overall search level will be combined into a general level.

After combining the similar questions, based on the search level of the visual analysis, and also the complexity of the user task, the generic questions are organized in increasing order, from elementary level to general level.

Organizing the questions will benefit the analysis step by making the analysis orderly and saving time on the visual analysis, and make it convenient to summarize the answers to the generic questions.

Research question 2 will be answered in this step.

3.3.3. Find gaps of the generic questions

The generic questions proposed may have gaps in the different aspects of the three components (what, when and where), which means the three components may not be equally dispersed among the proposed questions. In order to find those gaps, the generic questions were listed in one column of a table, with the three components in the same table, and using the number "1" for "not null".

3.3.4. Select visual outputs

Based on the two selected clustering methods and the case study big spatiotemporal dataset, from a set of corresponding visual outputs, the visual outputs that can reflect the characteristics of the spatiotemporal big data were selected.

Research question 4 will be answered in this step.

3.3.5. Visual outputs analysis

When the generic questions are prepared, the visual outputs analysis starts. A contingency table (a table is made up of rows and columns) will be formed based on the visual outputs and the generic question. And then look for the answers to the generic questions in visual outputs and fill in the table.

Research questions 5 will be answered in this step.

3.3.6. Formulate hypotheses

Summarize and analyze the answers to the generic questions, then formulate a series of hypotheses about what aspects of questions, what search level of questions these clustering visual outputs are able to answer.

Research questions 6 will be answered in this step.

3.3.7. Get specific domain questions

Once hypotheses are formulated, they need to be tested. The hypotheses can be tested by using a case study dataset to obtain the real information needs, and then to see if the answers to the real information needs are consistent with the answers to the generic questions (the two sets of answers are based on the same given visual outputs). This requires that the generic questions are compatible with the real information needs, but first, we need to know what the real information needs are.

For a specific dataset, initially domain experts (data processing and analyzing) have some information needs; these information needs can be formed into specific domain questions. To get to know the needed information, domain experts usually use clustering techniques. The visual outputs of clustering techniques will represent the dataset information. Therefore, in this research study, based on the case study dataset,

domain experts who have been working on data processing and analyzing field will be interviewed, to get the specific domain questions (real information need).
The domain questions are also generated in the same way as the generic questions. (I.e. from the three components of what, when and where; and from the three information search levels)
Research question 7 will be answered in this step.

3.3.8. Organize Domain questions

Domain questions are combined in the same way as generic questions, and also arranged in order of increasing hierarchy, on the basis of the search level.
Research question 8 will be partly answered in this step.

3.3.9. Compatibility between generic questions and domain questions

Because the two sets of questions (domain questions and generic questions) are generated from persons with different backgrounds, we consider that there will probably be missing questions between them. Only by making sure the two sets of questions are compatible, can the generic questions be verified by the domain questions. In order to find the missing questions, both of the sets of questions will be analyzed, and relevant questions from the two sets will be linked together.
If the generic questions and the domain questions are properly proposed, the best situation will be when the two sets of questions match each other well; otherwise, the generic questions will be considered to be modified or the domain experts will be requested to generate more on the missing aspects of the questions.
Research questions 8 will be partly answered in this step.

3.3.10. Experiment

An experiment will be set up and executed to test the formulated hypotheses, and to collect performance information on the clustering visual outputs. To achieve these purposes, suitable usability research methods are needed.

3.3.10.1. Introduction of usability test

Usability is about adaptability, ease-of use and learnability of a product for the users. The product can be a map, tool, software, website interface or any other man-made object that can be used by people to interact. According to ISO DIS 9241-11 (Guidance of Usability 1998), the definition of usability is:
"Usability is the extent to which a product can be used by specified user to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." The effectiveness is about the accuracy and completeness of achieving a certain goal; efficiency refers to how quickly a task can be accomplished; satisfaction is about how pleasant users feel about a certain design. (Frøkjær et al., 2000; "Usability Design and Evaluation," 2006)

Usability testing refers to testing a product with experiments using representative users. It can reflect the users' reactions to products by means of observing and analyzing their actual behaviours, and from the results of experiments from the execution of a predefined task.

Usability testing includes the main methods of heuristic evaluation, heuristic estimation, cognitive walkthroughs, formal usability inspections, pluralistic walkthroughs, feature inspection, consistency inspection, and standards inspection (Nielsen, 1994). Since then, many other new methods have been developed, including task analysis. There is/are usability technique(s) embedded in each method. Amongst these techniques, the commonly used and important ones are interview, questionnaire, think aloud, eye tracking, observation, and focus group (Razeghi, 2010).

3.3.10.2. Usability technique

The interview is a method for discovering information held by potential users of the object. There are two types of interview, structured and unstructured, which means with and without predefined questions. The unstructured interviews can gather lots of unexpected information, while in the structured interviews, the information collected is also structured and can target the needs, so that the results would be comparable. This is also the main characteristic of questionnaires (van Elzakker et al., 2004).

Questionnaires and interviews can be used to collect information without asking subjects to execute a task. However, in complex exploratory qualitative research, a clear and concrete task would be necessary to find answers to the research questions and meet the research objectives (van Elzakker et al., 2004). A detailed designed problem–solving task brings representative users and investigators together, so investigators can directly collect experiment data. During the task execution, observation and think aloud techniques are often used. Observation is to observe user's behaviours when they are executing a task, or to observe video recordings of users' behaviours of their task execution. And the think aloud technique requires test users to speak out their thoughts as they are executing a problem-solving task (Razeghi, 2010).

A focus group was originally called "focused interviews", it's a group of test users whose opinions are requested about a specific topic. It is an informal technique to obtain needs and feelings and ideas from a group of users (Nielsen, 1997).

"Focus groups are not generally appropriate for evaluation, but they bring individuals come together and express diverse views on the topic: useful not only to find the range of views, but also for the participants to learn from each other, and to generate a sense of social cohesion"("Focus groups,"). The investigator should act as a moderator who maintains the group's focus. Specific topics need to be preplanned, and the type of information to be collected needs to be predetermined. The moderator has the mission of keeping the discussion on track without limiting the ideas and comments, ensuring that every participant contributes to the group discussion, and preventing opinions from one person to be the domination.

Table 3.1 shows the main usability techniques and their investigator tasks.

Table 3.1 Main usability techniques(Razeghi, 2010)

Usability testing technique	Task of investigator
Interview	To elicits opinions held by potential users of the object being designed
Questionnaire	To do a survey or statistical study from answers for a set of questions
Think aloud	To analyze recorded verbal data from asking subjects to voice their thoughts when executing a problem solving task
Eye tracking	To capture eye movements when subjects are executing a task
Observation	To observe the execution of a task
Focus group	To elicit perceptions, feelings, attitudes, and opinion of subjects about the topic

Suitable usability techniques need to be found and adapted to this research study, so in this research study, we prefer to call them user research methods.

3.3.10.3. Select user research methods

The selection of user research methods is information requirements oriented. This means the selection should depend on what kind of data the research needs and what purpose the research is aiming to achieve. The selected methods should reflect what the real tasks are like, and should be able to answer the relevant research questions.

Generally speaking, there are two types of data from tasks, process data and bottom-line data. The former is from objective observations of what the test users do step by step and thoughts generated during completing tasks. From this type of data, people want to find the reason for the test users' behaviour and thinking; this is the focus of the think aloud method, while the latter gives a summary of what happened, it tells how good the design is, for instance, how long did users take, how many errors they made(Clayton Lewis, 1993).

According to the objectives of this research study, what the test users do and think while completing tasks needs to be known, and the reasons for the test users' decisions and thoughts need to be discovered; and how good, how fast and how successful the visual outputs perform in answering the questions need to be evaluated. Thus, this research is aiming at collecting insights and then is regarded as a qualitative research; both process data and bottom-line data are needed. Based on the above, we have decided on the corresponding user research methods:

Use test users to implement information-seeking task for testing hypotheses (I.e. This means test users search the answers to the specific domain questions in the given visual outputs.).

In addition to that, think aloud, observation and recording, and questionnaire are selected for getting to know the performance of the visual outputs. For this qualitative research, after the above methods, we also employ the focus group to explain the behaviour and decisions of test participants.

Research questions 9, 10, and11 will be answer in this part.

3.3.11. Conclusion

The proposed approaches are referring to and supposed to solve all the research questions. We expect to achieve the research objectives eventually by going through the whole process.

4. CASE STUDY

4.1. Case study description

A big spatiotemporal data was selected as case study dataset, which is the Netherlands twenty-year (from 1992 to 2011) daily temperature data from twenty-eight Dutch meteorological stations. Based on this real dataset, the visual outputs from SOM and co-clustering are selected, and the real information needs (domain questions) from domain experts are collected. The visual outputs, on the one hand are used for checking what kind of generic questions can be answered by them; on the other hand are used for testing what kind of domain questions they can answer. The domain questions have a connection to the generic questions; hence, the results of the generic questions can be verified by the results of the domain question by checking the consistency of the results.

4.2. Case study visual outputs selection

This dataset had been applied to the selected two clustering methods and has two corresponding types of clustering visual outputs products. The clustering method of SOM organize different aspects of information of data separately; thus, for this case study dataset, there are two sets of SOM visual outputs, based on the time information and the spatial information, called year-based SOM and station-based SOM respectively. The co-clustering groups the two aspects of information concurrently. As be mentioned in Chapter 2, the common visual representations of SOM are: u-matrix map, cluster map, geographic map, trend plot, and anomaly graph; and of the representations of co-clustering are heat map and geographical map. Out of all the visual representations, for this research study, cluster map, geographical map and trend plot from SOM visual outputs, and geographical map for co-clustering. The reasons for the selection are:

a. The selected visual outputs should be typical and represent the characteristics of clustering technique (show the trend and pattern),
b. The selected visual outputs should contain the typical information of the case study spatiotemporal data (spatial and temporal).
c. The selected visual outputs should contain more information than the rest of the representations.

Below are the selected clustering visual outputs.

1. Year-based SOM

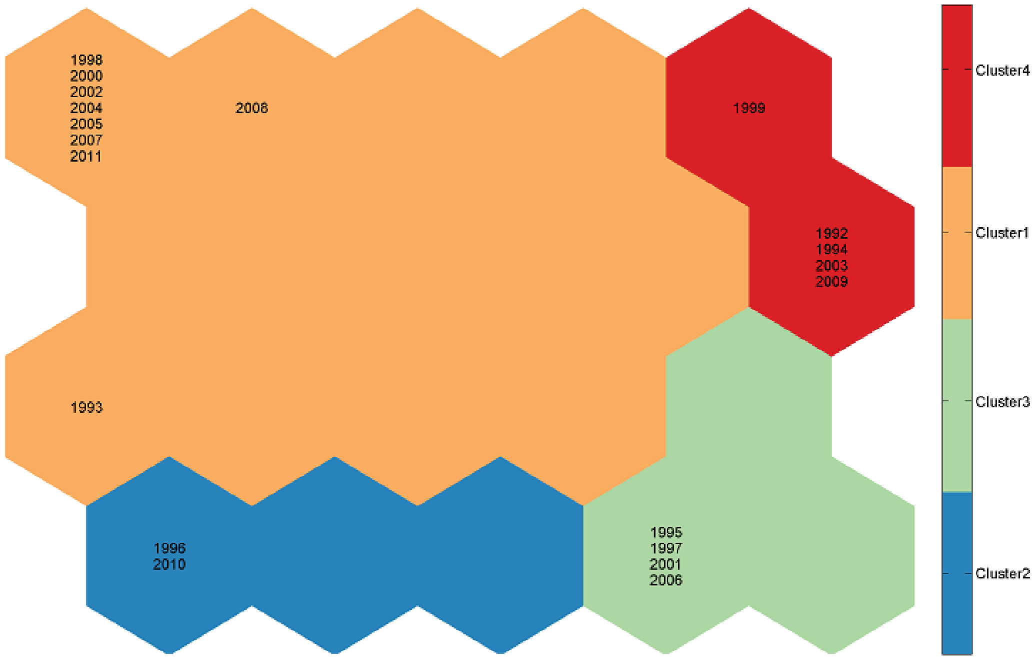


Figure 4.1 Year-based SOM cluster map

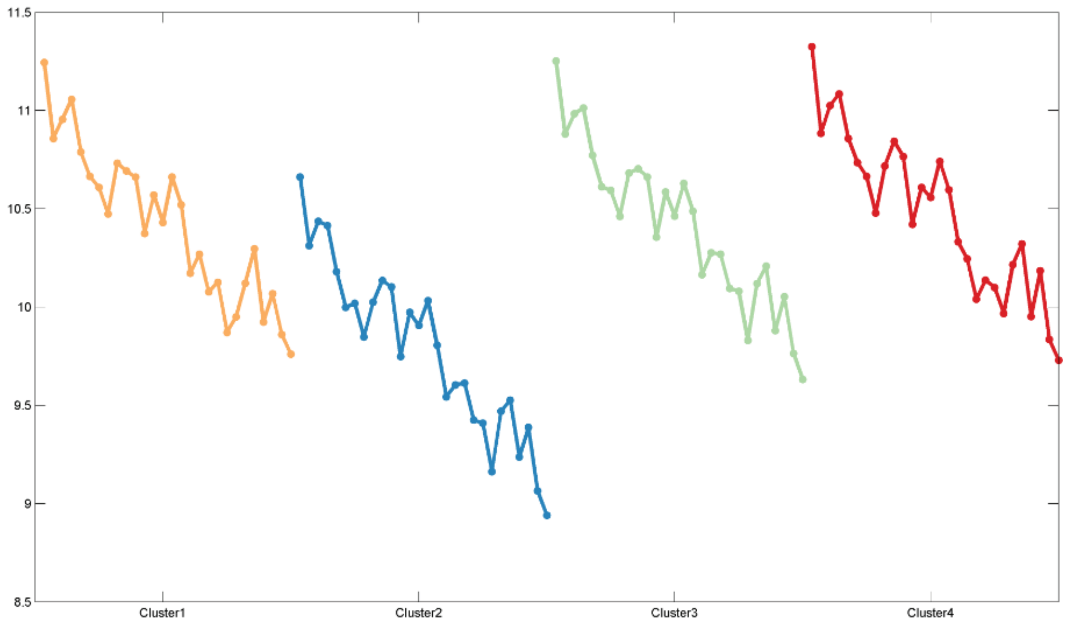


Figure 4.2Year-based SOM trend plot

Descriptions of Year-based SOM graphics:

Figure 4.1 is produced based on the similarity of 20-year daily temperature data. The four colours show four year-clusters. The colour scale from the top to the bottom indicates the attribute value from high to low. This graphic is a direct clustering result of time perspective.

Figure 4.2 is designed based on the year clustering result. The four lines are four year clusters. The 28 points on each line from left to right are the 28 meteorology stations from the southwest to the northeast of the Netherlands. The average annual values of temperature of each station are calculated. The numbers on the Y-axis of the diagram are the annual average values of the temperature. The year clusters on the x-axis and colours of the four lines are consistent with the cluster map.

As shown on the graphic, the 28 points on each line, from left to right, Station IDs (310, 319, 323, 330, 344, 210, 350, 348, 240, 380, 370, 356, 260, 375, 391, 235, 275, 269, 267, 273, 278, 270, 283, 277, 279, 290, 280, and 286).

2. Station-based SOM

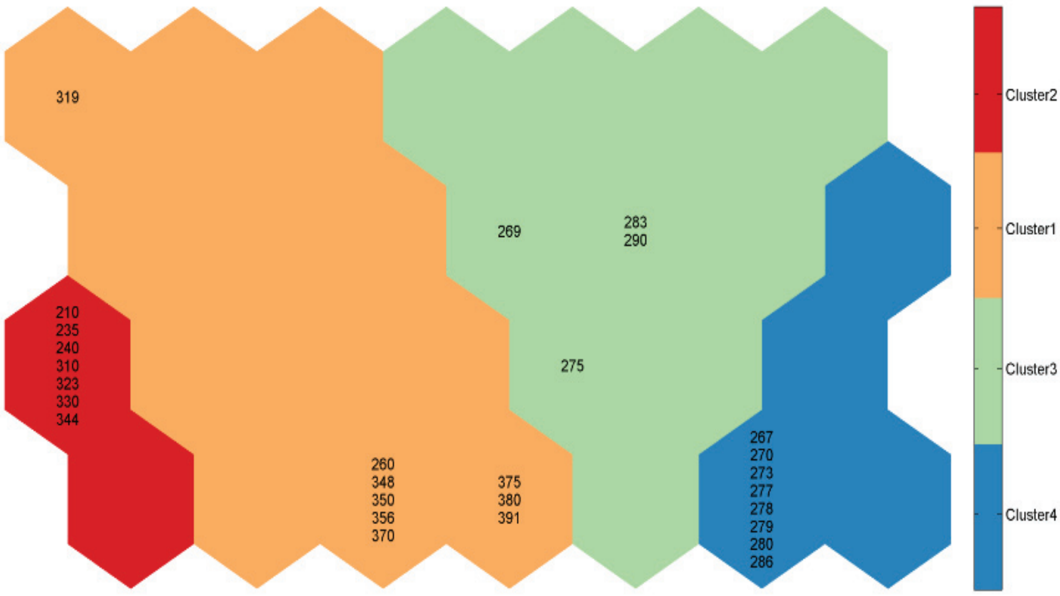


Figure 4.3 Station-based SOM cluster map

Figure 4.4 Station-based SOM geographical map

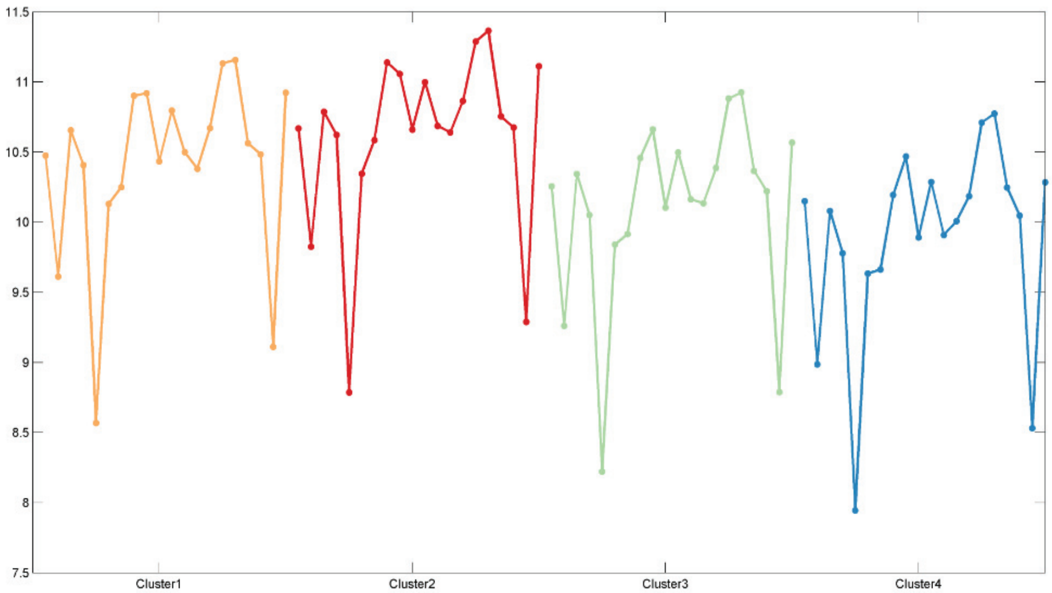


Figure 4.5 Station-based SOM trend plot

Descriptions of Station-based SOM graphics:
Figure 4.3 is produced based on the similarity of 20-year daily temperature data of all the stations. The four colours show four station-clusters. The colour scale from the top to the bottom indicates the attribute value from high to low. This graphic is directly from the clustering result of spatial perspective.
Figure 4.4 displays the information of cluster map on a geographic map.
Figure 4.5 is designed based on the station clustering result. The four lines are the four station clusters. In each station cluster, there are 20 points, which indicate 20 years; the average annual value of temperature of each year from the stations, between 1992 and 2011 are calculated. The numbers on the Y-axis are the annual average values of the temperature. The clusters on the x-axis and colours of the four lines are consistent with the cluster map. As shown on the graphic, the 20 points on each line, from left to right, are the 20 years: 1992-2011.

3. Co-clustering

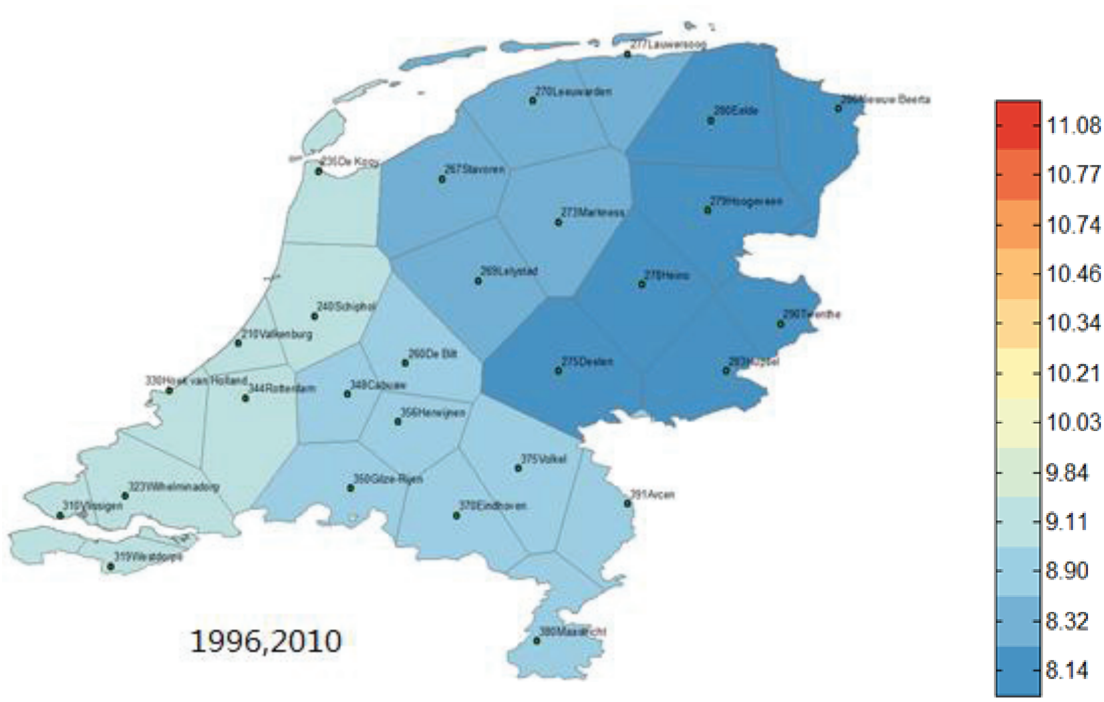


Figure 4.6 Co-clustering graphical map

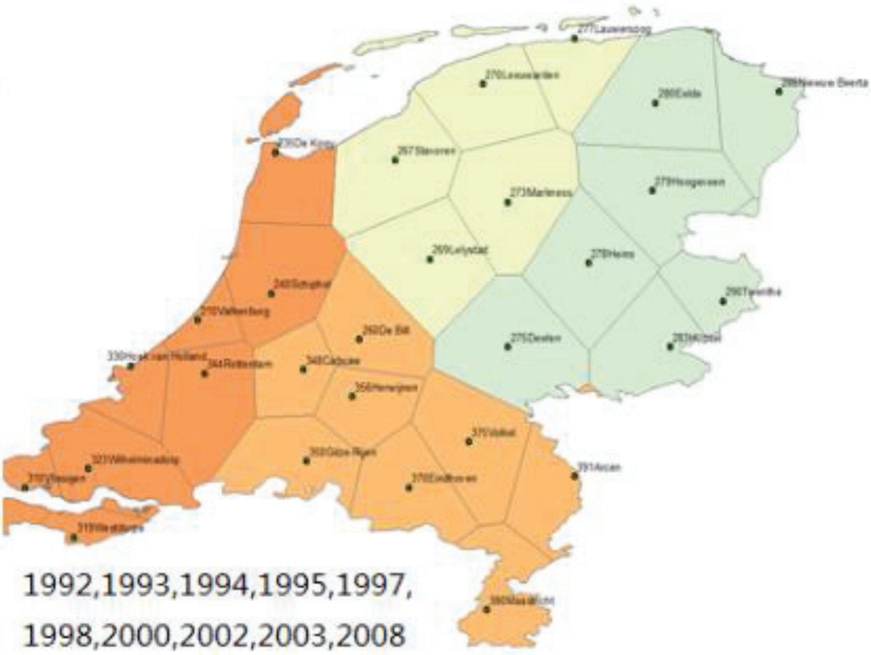


Figure 4.7 Co-clustering graphical map

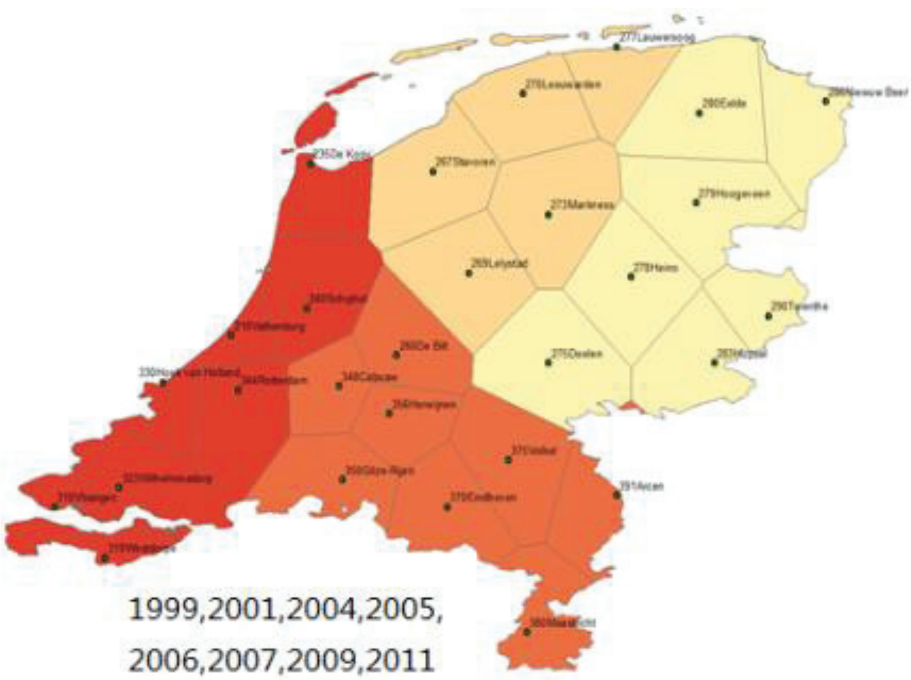


Figure 4.8 Co-clustering graphical map

Descriptions of co-clustering graphics:
The three graphics together represent the similarity of years and stations simultaneously.
Each map represents a cluster of years. Inside each year-cluster, the different colours show the different station-clusters. The colours with numbers on the colour scale indicate the average annual temperatures.

5. VISUAL ANALYTICS

5.1. Introduction

In (Keim et al., 2008), the authors described the goal of Visual Analytics as "to make our way of processing data and information transparent for an analytic discourse." They also offered a specific definition: "Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets." Visual analytics builds on a variety of scientific fields. Data mining is one of the main fields that benefit from it for deriving information from data analysis algorithm (in this research it is clustering algorithms) result -visual outputs.

Many visual analytic process designs are motivated by the so-called Visual Information-Seeking Mantra- "Overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996). The mantra can be adjusted and used in this research: first examine the representations from the whole space, time and attribute perspectives; second, focus on interesting patterns to analyze; third, inside the interesting patterns from the previous step, target details from different perspectives.

For the sake of analyzing and retrieving information from visual outputs, we need to know the information needs for the data. These needs can be formed into questions. Through finding the answers to these questions in clustering visual outputs, the analysis will be done. Proper questions will be the key factor which influences the analysis result. In order to draw conclusions about the analysis results, make it generic, we also need a typology to summarize the analysts' questions. Thus, the challenges are how to pose proper questions for analysis and how to summarize the proposed questions.

A schema will be helpful for proposing questions and making sure the questions cover different levels and aspects. Bertin (1983) proposed a systematic approach for arbitrary data, in which the author suggested a framework based on the "question types" and "read levels". He determined that there are as many question types as components (variables) in information. Inside each question type, there are three reading levels, overall, intermediate and elementary, which correspond to whether the questions are about all the objects of the data, a group of objects or a single object. However, the real problems could be much more complex than Bertin's question type. There could be different reading levels under the same question types and reading levels. So after Bertin's approach, Peuquet (1994) developed a more specific scheme for spatiotemporal data, and according to the three components (what, when and where), she proposed three basic possible questions:

1. When + where--what: Describe the objects or set of objects that are present at a given location or set of locations at a given time or set of times.
2. When+ what--where: Describe the location or set of locations occupied by a given object or set of objects at a given time or set of times.
3. Where+ what--when: Describe the times or set of times that a given object or set of objects occupied a given location or set of locations.

This categorization of question types also involves a similar notion as Bertin's reading level. This research would like to follow the same idea as (Andrienko et al., 2003), and use the term of "searching level" instead of “reading level”, mentioning the range of searching information in a single element, a group of elements or all the elements as a whole.

In the perspective of exploring the ability of the clustering visual outputs, we intend to find what kind of tasks the two clustering method visual outputs are able to support (i.e., to find what kind of questions about the spatiotemporal time series data the clustering visual outputs are able to answer). Bertin’s and Peuquet’s typology directly link the task to data and that is convenient for analysts. (Andrienko et al., 2003) extended Bertin’s classification scheme and simplified the scheme by “reducing the number of categories in each dimension”. They reduced the categories of “search level” by combining intermediate level and overall level and named the new categorization “general level” because they found that the exploration of an intermediate search level usually doesn’t vary much from exploring an overall search level. This approach was also applied in this research.

5.2. Generate and organize generic questions

The complexity of generic questions is in the aspects of search target and search range. For instance, the lower complexity level questions can be targeting only in the search range of when, and the more complex search range level could be both in when and where; even if the questions are in the same search range, the search target could be different in searching for a single element or multiple elements

In order to generate structured generic questions and make sure the generic questions cover all the main aspects, we follow the scheme that proposed by Peuquet (1994), which has been mentioned in the last section. As we discussed in the Chapter 3, the generic questions are organized in increasing order; the similar generic questions that differ in the intermediate level and overall level are combined together; and the generic questions are marked with number “1” on the involved components, for finding gaps in the components and levels.

The final generic questions and organized results are shown in Table 5.1.

This table shows that:

1. The generated generic questions are distributed properly on the three components of what, when and where.
2. There are only a few elementary questions, and more general questions. This is because the exploration of big data is not for information query, but for finding the pattern and trend. And pattern and trend will be shown in a larger time or spatial range.

Table 5.1 Generic questions

Number	Generic questions	What	When	Where
1	What is the attribute value for a given year in a given place?	1		
2	Can the highest (lowest) attribute value of a given place of a given year be found?	1		
3	What is the highest (lowest) attribute value of a given place?	1		
4	Which places have similar attribute value of the whole study area in a given year?			1
5	Which places are similar?	1		1
6	Which years are similar?	1	1	
7	Where has the highest (lowest) attribute value in a given year?	1		1
8	Where has (have) the highest attribute value in the whole study area over the whole time period?	1		1
9	When did the highest (lowest) attribute value happen in a given place?	1	1	
10	Which year(s) have the highest (lowest) attribute value over the whole time period?	1	1	
11	How is the change of the attribute value in a given place?	1	1	
12	Where has (have) the highest (lowest) attribute value in a given area during a time period?		1	1
13	Which year (years) has (have) the highest (lowest) attribute value in a given area?	1	1	1
14	Where has (have) the highest (lowest) attribute value during a time period?	1	1	1
15	What is the highest (lowest) attribute value in the whole study area over the whole time period?	1	1	1
16	What's the trend of the attribute in the whole study area over years?	1	1	1

5.3. Analysis result of the selected three types of visual outputs

The three selected types of visual outputs were used to answer generic questions separately, "Y" indicates that the question could be answered by the corresponding visual outputs, and "N" indicates that the question couldn't be answered by the relevant visual outputs.

Year –based SOM visual analysis result

Table 5.2 Generic questions year-based SOM visual analysis result

Num	Generic questions	Year based SOM	
		Cluster map	Trend plot
1	What is the attribute value for a given year in a given place?	N	Y
2	Can the highest (lowest) attribute value of a given place of a given year be found?	N	N
3	What is the highest (lowest) attribute value of a given place?	N	Y
4	Which places have similar attribute value of the whole study area in a given year?	N	N
5	Which places are similar?	N	N
6	Which years are similar?	Y	Y
7	Where has the highest (lowest) attribute value in a given year?	N	Y
8	Where has (have) the highest attribute value in the whole study area over the whole time period?	N	Y
9	When did the highest (lowest) attribute value happen in a given place?	N	Y
10	Which year(s) have the highest (lowest) attribute value over the whole time period?	Y	Y
11	How is the change of the attribute value in a given place?	N	N
12	Where has the highest (lowest) attribute value in a given area during a time period?	N	N
13	Which year (years) has (have) the highest (lowest) attribute value in a given area?	N	N
14	Where has (have) the highest (lowest) attribute value during a time period?	N	Y
15	What is the highest (lowest) attribute value in the whole study area over the whole time period?	N	Y
16	What's the trend of the attribute in the whole study area over years?	N	N

Station-based SOM visual analysis result

Table 5.3 Generic questions station-based SOM visual analysis result

Num	Generic questions	SOM Station based		
		Cluster map	Geo-map	Trend plot
1	What is the attribute value for a given year in a given place?	N	N	Y
2	Can the highest (lowest) attribute value of a given place of a given year be found?	N	N	N
3	What is the highest (lowest) attribute value of a given place?	N	N	Y
4	Which places have similar attribute value of the whole study area in a given year?	N	N	Y
5	Which places are similar?	Y	Y	Y
6	Which years are similar?	N	N	N
7	Where has the highest (lowest) attribute value in a given year?	N	N	Y
8	Where has (have) the highest attribute value in the whole study area over the whole time period?	Y	Y	Y
9	When did the highest (lowest) attribute value happen in a given place?	N	N	Y
10	Which year(s) have the highest (lowest) attribute value over the whole time period?	N	N	Y
11	How is the change of the attribute value in a given place?	N	N	Y
12	Where has the highest (lowest) attribute value in a given area during a time period?	N	N	N
13	Which year (years) has (have) the highest (lowest) attribute value in a given area?	N	N	Y
14	Where has (have) the highest (lowest) attribute value during a time period?	N	N	Y
15	What is the highest (lowest) attribute value in the whole study area over the whole time period?	N	N	Y
16	What's the trend of the attribute in the whole study area over years?	N	N	Y

Co-clustering visual analysis result

Table 5.4 Generic questions co-clustering visual analysis result

Num	Generic questions	Co-clustering geo-map
1	What is the attribute value for a given year in a given place?	Y
2	Can the highest (lowest) attribute value of a given place of a given year be found?	N
3	What is the highest (lowest) attribute value of a given place?	Y
4	Which places have similar attribute value of the whole study area in a given year?	Y
5	Which places are similar?	Y
6	Which years are similar?	Y
7	Where has the highest (lowest) attribute value in a given year?	Y
8	Where has (have) the highest attribute value in the whole study area over the whole time period?	Y
9	When did the highest (lowest) attribute value happen in a given place?	Y
10	Which year(s) have the highest (lowest) attribute value over the whole time period?	Y
11	How is the change of the attribute value in a given place?	N
12	Where has the highest (lowest) attribute value in a given area during a time period?	N
13	Which year (years) has (have) the highest (lowest) attribute value in a given area?	Y
14	Where has (have) the highest (lowest) attribute value during a time period?	Y
15	What is the highest (lowest) attribute value in the whole study area over the whole time period?	Y
16	What's the trend of the attribute in the whole study area over years?	N

5.4. Summarize the result of visual analysis and Formulate hypotheses

5.4.1. Year-based SOM

From the clustering literature review, a nature can be seen that the similar objects are in the same cluster. The cluster map clearly showed years clusters and the legend of colour scale showed the attribute level of each cluster, these information easily answered questions 6 and 10.

The trend plot answered more questions than the cluster map, including the two questions answered by the cluster map. It showed the average annual attribute value of each station, with each point in each year-cluster.

In summary, the results show the SOM year-based focus on the temporal dimension, the hypotheses are: the selected year-based SOM visual outputs can answer most of the questions about attribute, location, and time when the searching object or ranges are specified. And it can answer different stations information inside each year cluster. It cannot, however, answer questions about searching for the extreme value for a specific place in a given year (question 2), similar stations (question 4 and 5), geographical information (12 and 13), and the attribute values change over years (question 11and 16).

5.4.2. SOM station-based

The SOM station-based cluster map showed station cluster and the colour scale showed the attribute value levels of the clusters, these information easily answered questions 5 and 8.

The geographical map contains all the information from the cluster map, in addition to geographical information.

The trend plot offered the most information compare to the cluster map and geographic map. It showed the average annual attribute value for each year, with each point in every station-cluster. And because all the years are arranged in the time-change order, this graphic can also answer time-series questions.

In summary, the results show the SOM station-based focus on the spatial dimension. The hypotheses are: the selected station-based SOM visual outputs can answer most of the questions about attribute, location, and time when the searching object or searching ranges are specified. And it can represent trend information of attribute change over years inside each station cluster. It cannot answer questions about the extreme values for specific locations in a given year (question 2), similar years (question 6), and the location differences in a station cluster (question 12).

5.4.3. Co-clustering

The co-clustering visual output is a combination of three single geographical maps. It offered similarity information of both time and location. And the colour scale indicated different attribute levels.

From the analysis, the hypotheses are: the selected co-clustering visual outputs can answer most of the questions about attribute, location, and time when the searching object or searching ranges are specified. It cannot answer about the extreme values for specific locations in a given year (question 2), attribute change over time (question 11 and 16), and cannot find the different location inside a station cluster (question 12). All the attribute information offered from this selected co-clustering visual output is about the attribute of time or station cluster, which means it only compare attribute information between clusters.

5.4.4. Summary of the visual analysis

The different types of graphics have their own advantages, and the visual outputs of year-based and station-based SOM compensate for each other. To answer the same questions, it's easier to use co-clustering visual outputs; if the exact information about a year or a station is required, instead of a year-cluster or a station-cluster, then the station-based and year-based SOM should be used, respectively (for example, questions 7, 8, 9,and 14). The three ways of visual outputs all cannot answer questions 2 and 12, because clustering result doesn't offer the specific daily attribute information (question 2) and doesn't distinguish a different year in a year cluster or a different station in a station cluster.

5.5. Domain questions

The real information need is collected by interviewing domain experts.

The interview phase includes two steps: first, request an appointment to domain experts. During the communication, this research study was simply introduced to them as well as the expectation for them. To generate proper specific domain questions will take their quite a time. Thus, in order to make this process more efficient, an email (Appendix 1: interview email) with detailed information about the case study dataset and the requirements was sent to the domain questions. The requirements of the domain questions are based on the same considerations as generic questions: the questions should cover different levels and all the main aspects (components) of the spatiotemporal dataset. The generated domain questions are also organized in increasing hierarchical order on the basis of question complexity, and the similar questions are combined together.

The result domain questions are listed below:

Table 5.5 Organized domain questions

Organized domain questions	
1	What is the lowest temperature in 290 in 2010?
2	What is the lowest temperature in 290 throughout the whole study period?
3	Is there any other place in the Netherlands as cold/warm as 290 in 2010?
4	Is there any other place generally as cold/warm as 290?
5	Is there any other year as cold/warm as 2010 in the Netherlands?
6	Where is the coldest place in 2010?
7	When is the coldest year in 290?
8	Which year(s) is/are the hottest throughout the period in the Netherlands?
9	What is the pattern of change of temperature in 290 throughout the whole study period?
10	Where is the coldest place in the Southwest of the Netherlands from 2005-2008?
11	Which year(s) is (are) the hottest year between 2000 and 2005 in the northeast of the Netherlands?
12	Where is (are) the hottest place between 2000 and 2005 in the whole Netherlands?
13	What is the lowest temperature in the Netherlands during all years?
14	What is the pattern of change of temperature in the whole Netherlands?

5.6. Check compatibility

Through analyzing and comparing two sets of questions, the corresponding questions were linked together, and then the missing questions were found. See Table5.6.

Table 5.6 Compatibility between generic questions and domain questions

Generic questions	Domain questions
1. What is the attribute value for a given year in a given place?	
2. Can the highest (lowest) attribute value of a given place of a given year be found?	1. What is the lowest temperature in 290 in 2010?
3. What is the highest (lowest) attribute value of a given place?	2. What is the lowest temperature in 290 throughout the whole study area?
4. Which places have similar attribute value of the whole study area in a given year?	3. Is there any other place in the Netherlands as cold/warm as 290 in 2010?
5. Which places are similar?	4. Is there any other place generally as cold/warm as 290?
6. Which years are similar?	5. Is there any other year as cold/warm as 2010 in the Netherlands?
7. Where has the highest (lowest) attribute value in a given year?	6. Where is the coldest place in 2010?
8. Where has (have) highest attribute value in the whole study area over the whole time period?	
9. When did the highest (lowest) attribute value happen in a given place?	7. When is the coldest year in 290?
10. Which year(s) have the highest (lowest) attribute value over the whole time period?	8. Which year(s) is/are the hottest throughout the period in the Netherlands?
11. How is the change of the attribute value in a given place?	9. What is the pattern of change of temperature in 290 throughout the whole study period?
12. Where has the highest (lowest) attribute value in a given area during a time period?	10. Where is the coldest place in the Southeast of the Netherlands from 2005-2008?
13. Which year (years) has (have) the highest (lowest) attribute value in a given area?	11. Which year(s) is (are) the hottest between 2000 and 2005 in the northeast of the Netherlands?
14. Where has (have) the highest (lowest) attribute value during a time period?	12. Where is (are) the hottest place(s) between 2000 and 2005 in the whole Netherlands?
15. What is the highest (lowest) attribute value in the whole study area over the whole time period?	13. What is the lowest temperature in the Netherlands during all years?
16. What's the trend (spatial and temporal perspective) of the attribute in the whole study area over years?	14. What is the pattern of change of temperature in the whole Netherlands?

The table shows there are two generic questions without corresponding domain questions. The two generic questions are quite straightforward for analyst to find the answers, so this will not impact the research moving on to the next step. This result illustrates that the two sets of questions match each other well.

5.7. Conclusion

In this chapter, based on the generic questions, the clustering visual outputs were analyzed the selected, and a series of hypotheses were formulated by integrating the visual analysis results. And then the compatibility between generic questions and the domain questions were verified, which confirmed the generic questions are qualified and the domain questions can be used in experiment to test the hypotheses.

6. USER RESEARCH IMPLEMENTATION

6.1. Overview

In this research study, an experiment was implemented in order to test the formulated hypotheses and describe the satisfaction, efficiency and effectiveness of the three selected types of visual outputs. Three sets of visual outputs were presented to individual test persons; with the help of the visual outputs, test users answered specific domain questions, executed think aloud tasks, and completed questionnaires, while the investigator used several user research methods to obtain the data from the experiment. This chapter describes the setup and the implementation of the experiment, which includes the selection of test participants, a detailed plan and the implementation, along with a critical analysis of the implementation process. Finally the experiment results were presented, analyzed and interpreted.

6.2. Test participant

6.2.1. Selection of test participants

"The point of testing is to anticipate what will happen when real users start using your system. So the best test users will be people who are representative of the people you expect to have as users"(Clayton Lewis, 1993). Participant recruiting is, next to the selection of the evaluation technique, one of the most important and basic steps in all usability studies. But due to time or budget constraint, this critical component is not always high enough on the priority list. Without right test users, proper results cannot be obtained (NIELSEN, 2003).

To know what kind of people are real users, it is needed to know in what field the research result will be used. Often, it is difficult to find those real users, and people who are more or less related to the research will be taken into consideration.

Being a test user can be distressing, and some people may feel embarrassed with audio/video recording that is created while they are completing a task or even, if just their name is linked with test results/performance. To avoid these issues, the participants should be voluntary, informed that their name will not be revealed. This means the participants are willing to join user tests and are informed and agreed with everything about the experiments before they are starting.

While choosing the right test users one also needs to be careful about inviting friends, colleagues or some other people who are familiar with the investigator, because they may not feel free to decline. If they would like to join, make sure they have no hesitation and are really interested in getting the task done.
The number of test participants should depend on the nature of the research (quantitative or qualitative), and what tasks will be involved in the experiments. As discussed in chapter 3, this research is a qualitative research because it is aiming at collecting insights.

This research study is motivated by obtaining information from spatiotemporal big data, by means of analyzing clustering visual outputs. Thus, the real potential users, on the one hand, would be people who require information from this kind of big data; on the other hand, they should be able to analyze visual representations and are interested in being engaged in the test. Due to time limitations, it is not possible to

allocate resources in finding real users (the ones who use clustering technique to analyze a spatiotemporal big dataset). Therefore student researchers and teachers in several departments of ITC were considered to be the test users, since some of them are doing research relevant to spatiotemporal big data.
In a qualitative research, "Test 5 users in a usability study. This lets you find almost as many usability problems as you'd find using many more test participants." (NIELSEN, 2012). But there are also some exceptions for this minimum number of test person rule, like focus group; it requires a minimum of six test users ("Focus groups,"). In (NIELSEN, 2000), the author proposes that "Even three users are enough to get an idea of the diversity in user behaviour and insight into what's unique and what can be generalized." In this study, the three ways of clustering visual outputs were arranged to be tested in three different orderings; hence, for each group, the minimum three users were needed for each order and nine test users were used in total to execute the information-seeking task.

6.2.2. Invite test participants

Nine test persons were recruited via an invitation email (Appendix 2), considering about the above constraints. They are MSc or PHD researchers from different departments of ITC, working with spatiotemporal dataset.

6.3. Test site

Test environment is another important factor that ensures the quality of the user research. In this research, the ITC usability laboratory was chose as the experiment site, as it is specially suited for the execution of experiments, and has a video recorder, laptop and a quiet environment that are needed.

6.4. Experiment with individual test persons

6.4.1. Introduction

This stage of the experiment involved nine participants performing problem-solving tasks (information-seeking task) and think aloud simultaneously, and then completing a questionnaire; in the meanwhile, evaluation methods such as observation and recording have been applied by the investigator. This section includes the detailed description of the tasks and methods, the plan, the implementation and the results of the test.

6.4.2. Experiment description

Participants seek information in the selected three ways of clustering visual outputs to answer the specific domain questions (chapter 5, Table 5.5) and fill in the answers on the question form. While performing the information-seeking task, these persons were asked to continuously vocalize their thoughts specific to the task. With the results of the tasks, the hypotheses that were created from the generic questions could be tested.

During the think aloud task execution, the test participants needed to say what they were trying to achieve, what questions arose and what things they were reading. Their comments and expressions, and the time consumed when solving each question were recorded, and observed to make sure the observer can tell what test users were doing and where their comments fit into the sequence.

For recording, taking notes in summary form could also be practical. But in one respect, it takes some practice to do this fast enough to keep up in real time; in another respect, the mimics and the execution process for advanced analysis also need to be recorded, so the video recording was needed while the subject was executing the think aloud task (Clayton Lewis, 1993). In this instance, combining a video recording with written notes was suitable. During the task execution, the investigator was required to

remain silent so as to avoid biasing test persons' behaviour with untimely hints. Help to test users can only be given when the investigator receives no more useful information without intervention. If intervention occured and help was given, the process also had to be recorded.
After the execution of the information-seeking tasks, test users had to make choices on a Likert scale questionnaire (Appendix 4). This questionnaire was used for collecting opinions about the satisfaction of the users for each visual output selected.
The questions in the information-seeking task were the domain questions listed in the chapter 5, Table 5.5.

6.4.3. User research plan

The plan for the experiment with individual test persons involved three parts: before, during, and after the individual tests.
Before starting the individual user tests
Formed a checklist (Appendix 5) of the items for the experiment, prepare and remind the participants two days before starting the experiment.
Everything on the checklist were set up by the investigator (for example exhibiting the domain question form on the computer, putting the three hard copies of the visual outputs (Appendix 6) that are going to be tested on the table etc.), making sure everything needed is there and works properly.
In the first phase, test users were take part in the experiment separately at different times. As test participants arrival, they were first be warmly welcomed (Appendix 7), and then their basic characteristics were collected by a test participants information form (Appendix 8). The test participants information form mainly collect information in: education and work background, experience in clustering and in analyzing clustering visual outputs, and experience with analyzing spatiotemporal datasets.
The basic theory of the clustering technique, the think aloud method, descriptions and instructions of the tasks were explained by giving them "Introduction of clustering technique and the experiment" (Appendix 9).
The nine participants were randomly separated into three groups. They executed exactly the same tasks with the same domain questions, but the difference was the order in which they used visual outputs. That means that the visual outputs of the three ways of clustering were used by the three groups in different orders (the three orders are: co-clustering, year-based SOM, station-based SOM; year-based SOM, station-based SOM, co-clustering; and station-based SOM, co-clustering, year-based SOM). By removing ordering effects, we can have more confidence in the efficiency and effectiveness of the visual outputs. The search targets in the domain questions referring to different categories of visual outputs were slightly changed. By slightly changing the search target, the test users can be prevented from copying the answers from the other graphics.
Each test participant was given five minutes to become familiar with the visual outputs and then two domain questions to answer (appendix 3), as practice, before executing the real task. Basic explanations of the graphics, which were written on the same paper as the three hardcopies (Appendix 6-1, 6-2, 6-3), were given. This step was taken in order to help test users understand the task and the graphics. The practice domain questions are not drawn from the set of test domain questions, they are similar questions to the test domain questions that we removed from the set when organizing the domain questions (section 5.5).

During the individual user tests
Once the test commenced, participants were informed about the start of the video recording. In the process of the task execution, the following steps were taken:

1. The time the users spent on solving each question corresponding to the graphics was registered and filled in the time cost table.
2. Notes about what was happening during the execution and video recording were taken.
3. From other research experiences, during the task, saying nothing after the initial instructions usually won't work. Most people wouldn't give a good flow of comments without being pushed a bit. So the

test users were prompted to keep up the flow of comments and encouraged to talk.

4. Help wasn't offered during task, only when it was necessary, which means no more useful information will be received without intervention.
5. The laboratory environment was maintained, to keep the interruption away; and the equipments used for the test were made sure to work normally.

After the individual user tests

Test users will be requested to make their choice on the Likert scale questionnaire to collect their opinion about the visual representations.

6.4.4. Pilot study

Before the final evaluation experiment, a pilot study has been conducted. During the pilot test all steps as planned for the final test were executed to find potential weaknesses and errors. It had to reveal if the explanation and instruction to the test users were clear, if the individual steps in the plan were well connected, if the test questions were well understood and if the recording of think aloud went well. All the defects were modified before the real test. A pilot study is important to avoid trouble in a think aloud study. (Clayton Lewis, 1993)

Implementation of pilot study

Two ITC staff in GIP department joined in this pilot study. One is a usability expert and the other is a visual analysis expert. The two pilot studies were on Jan 22nd, in the morning and in the afternoon separately.

Everything was prepared and conducted according to the "user research plan" (6.5.3). The video was set in front of the test users, so that it could record not only the voice but also the mimics.

Several problems were discovered from the observation of the pilot test, and improvements have been made according to them. (Table 6.1)

Table 6.1 Problems discovered from the pilot study and corresponding improvements

Problems	Improvements
1.The visual outputs needed more explanations.	1. The details were added and separated to each graphic, and example graphics will be shown on PowerPoint.
2.The test explanation needed more details.	2. The details for the illustration about how to use each graphic- separately or as a combination will be added.
3.Questions 4 and 5 were confusing;	3. The wordings will be changed into: 4.which other station(s) has (have) the similar temperature as 310? 5. Which year(s) has (have) the similar temperature as 2010?
4.Write down answers cost lots of time	4. The answers were suggested to be written in a simple way.
5. "Introduction of the clustering theory and the first part of experiment" didn't impress users, and "speech to users" in the test didn't make the test smoothly.	5. Used "test scenario" (Appendix 10) and a "PowerPoint" (Appendix 11) instead of the two documents.
6. The question sheet for the three groups of visual outputs connected together led to confusion.	6. The question sheets were separated into three groups according to the three ways of clustering.
7. The questions answered by row wasted more time than answered by column.	7. The questions were answered by column based on the improvement 6 above.
8. The practice questions didn't cover all the items appeared in the question	8. A different type of question will be added: Is there any other place in the Netherlands as cold/warm as 319 in 2008?
9. A test user used the second visual outputs hard copy to answer questions in the first group.	9. The tests were interfered when seeing a test user making a mistake like this.

6.4.5. Implementation of the experiment with individual test persons

The time for the user test was arranged, considering the convenience of the test participants; final time schedule was sent to all the test users and each test user was informed one day before the test started. The individual user tests were conducted from January 28th to January 31. Each test user spent around 1.5 hours in all, during this time, on average approximately35 minutes were used for training. The nine test users were arranged as the table below.

Table 6.2Test persons and visual outputs arrangement

Group	**Test Person**	**Order of the tested visual outputs**		
Group 1	p1	co-clustering	Year based SOM	Station based SOM
	p2			
	p3			
Group 2	p4	Year based SOM	Station based SOM	co-clustering
	p5			
	p6			
Group 3	p7	Station based SOM	co-clustering	Year based SOM
	p8			
	p9			

People in the same group used the same order of question forms that refer to the visual outputs. Every step was the same for all the test participants.

Each video recording covered all the questions, from first to last, which means the training phases were not included. So the test users felt free to communicate with the test investigator and to be trained. As planned in the test scenario, for each participant, first the basic information was collected; second, a PowerPoint presentation containing basic theory relevant to the test and task introduction was shown to the user, and in the meanwhile, their questions were answered based on the contents of the presentation. Third, each test user was required to finish the three pre-prepared warm-up questions. The warm–up questions mention the temporal and spatial information, which helped the test users to get familiar with the terms in the questions and to link the terms to the graphics. From observation, we found there was an overloaded of information to the test users, who found it hard to remember and understand everything in a short time. That is why they used a rather long time to search for the answers to the warm-up questions.

During the video recording, for each question, the duration of time was registered between the starting and the answering of the question. On the question form, test users were required to fill in the form in this way: if they think the graphic can answer the corresponding question, write down the answer, otherwise fill in "N". During their thinking aloud, test users were reminded when they forgot to speak out their thoughts; they were requested to give their argumentation when they only gave the answer; and were asked to tell what they were considering when they said "it is confusing"; and were asked the reason when they said "the question is not clear".

After finishing each video recording, each test user answered Likert scale questionnaire based on their experience with the visual outputs in the test.

The domain question form answered by each test user was analyzed as soon as the test ended. The notes and the video recordings that were taken during the task executions were the bases for explaining the reasons to the answers; all the answers to the questions related to different visual outputs were analyzed, and the recordings were sufficient to explain the decisions users made during test. Therefore, no specific topics from the individual tests were found that needed to be brought to the focus group to discuss; only the choices from the Likert scale questionnaire were discussed in the focus group and the feedback about the test was asked.

6.5. Focus group

6.5.1. Introduction

In the second stage of the user research, the focus group method was applied. This section includes the description, plan and implementation of the focus group.

6.5.2. Description

The focus group was used in this experiment for discussing experiences of the test, explaining the decisions and obtaining suggestions about the visual outputs used.
The data that came from the focus group could be analyzed in detail, but since the participants' reactions and ideas were unstructured, detailed analysis appeared to be difficult and time consuming. Therefore, after the session, data analysis was formed a simple short report written up by the moderator, summing up the idea and views from the focus group (Nielsen, 1997)

6.5.3. Implementation for the focus group

The focus group involved all nine participants from the individual experiment. The site was the discussion room in the ITC hotel. The focus group was started on Feb 1st, at 10am, and test users were informed one day before the focus group started. The whole procedure lasted 65 minutes.

The reasons for the evaluations of the tested visual outputs they made on the Likert scale questionnaire were discussed. The hard copies of the visual outputs that were used on the test were given to them, as well as the likert scale questionnaires they completed. And then the questions on the Likert scale questionnaire were proposed one by one to ask the reasons for their decisions, based on their experience in individual experiments. The focus group also collected feedback on the individual test; all the test users thought the test was too intensive, which might cause more errors; and some of the domain questions could be understood in different ways, which consumed more time to answer. In the meanwhile, notes were taken to record their reasons and suggestions.

6.6. Implementation process analysis

The whole implementation was executed as improved plan. All the nine test participants understood the test introduction and illustration properly and went through the whole experiment. The plans went well; all the users spoke out their reasons for each question, and with the questions that were understood in different ways, test users all gave their reasons and corresponding answers, which offered sufficient information for analyzing the experiment result. But the training time took longer than expected, because the test users needed time to link the terms to the tested graphics and to understand the questions. Indeed, this is an indication that the "learning effect" plays a role, so that in the real experiment, the test users performed much faster than they did in the training phase to answer practice questions. However, because the questions are totally different, so are the levels of difficulty, thus the time taken between questions are hard to compare at the same level. This is the concern that this research study only compared the time taken between visual outputs on answering the same questions, and didn't compare the time difference between questions.

6.7. Implementation results and analysis

6.7.1. Analysis of the information-seeking task result

The understanding is the basic purpose for a qualitative data analysis, and the interpretation is the product of any qualitative analysis (Kaplan & Maxwell, 2005). In qualitative research, the inherence of flexibility and individual judgement reduce some of the reliability but the validity is often enhanced because the researcher's close attention to meaning, context, and process make it unlikely that important information will be skipped. Therefore, the shortfall of less reliability is compensated for by the strong validity that comes from the researcher's insight, cognitive ability and tacit knowledge (Kaplan & Maxwell, 2005).

During individual experiments, the domain question forms required test users to fill in the form with "Y" or "N" to indicate whether they could find the answer or not; with a "Y" answer, they were required to write down the exact answer. Some questions were understood by test users with different meanings, which means with one understanding the answer was Y, while with another understanding the answer was N. Test users gave their reasons and the corresponding answers for both interpretations. These situations were all recorded and were taken notes. The first step in analyzing the results was to check these double-answer questions from the video recording and the notes taken, and then write down the different answers and explanations for each this kind of question. All the answered question forms were compared; the strategies for their information-seeking were analyzed and interpreted.

The answers to the corresponding questions with the same "Y" or "N" and with the same validity explanations were regarded as the same and as correct answers. In cases in which the answer is interpreted as correct and agreed upon by the majority of the test users, the answer is regarded as a reference that is used to test the hypotheses; in cases in which the answer is regarded as correct but agreed upon by less than the majority of test persons, then further analysis is required. The indicator of a majority is 6. For each group of clustering visual outputs, the numbers of agreements on correctness to each domain question were counted and are shown in the tables below:

Table 6.3 Agreements on correctness of co-clustering

Agreements on correctness of co-clustering		
	Co-clustering	
Domain questions	**Geo-map**	**Agreed answer**
1. What is the lowest temperature in 286 in 2010?	5	N
2. What is the lowest temperature in 286 throughout the whole study period?	8	Y, Lowest mean temperature, answer is 8.14; single day, no answer
3. Is there any other place in the Netherlands as cold/warm as 286 in 2010?	9	Y, Objects in dark blue cluster
4. Is there any other place generally as cold/warm as 286?	9	Y, The stations in the same cluster as 286
5. Is there any other year as cold/warm as 2011 in the Netherlands?	9	Y, In the same year cluster
6. Where is the coldest place in 2011?	8	Y, Objects in light green cluster
7. When is the coldest year in 286?	9	Y, 1996;2010
8. Which year(s) is/are the coldest throughout the period in the Netherlands?	9	Y, 1996;2010
9. What is the pattern of change of temperature in 286 throughout the whole study period?	9	N (can't see time change pattern)
10. Where is the coldest place in the Southeast of the Netherlands from 2005-2008?	7	N (only one value)
11. Which year(s) is (are) the coldest year between 2000 and 2005 in the northeast of the Netherlands?	8	Y, 2001;2004;2005
12. Where is (are) the coldest place between 2000 and 2005 in the whole Netherlands?	8	Y, Objects in light green cluster
13. What is the coldest temperature in the Netherlands during all years?	9	Y, 8.14(exact value, no; year value, 8.14)
14. What is the pattern of change of temperature in the whole Netherlands?	9	Y, The temperature decreases from SW to NE of the Netherlands)

Table 6.4 Agreements on correctness of Year-based SOM

Agreements on correctness of Year-based SOM				
	Year based SOM			
Domain questions	**Cluster map**	**Agreed answer**	**Trend plot**	**Agreed answer**
1. What is the lowest temperature in 290 in 2010?	9	N	8	N
2. What is the lowest temperature in 290 throughout the whole study period?	9	N	9	(single day, no) Y, lowest mean, 9.3
3. Is/Are there any other place(s) in the Netherlands as cold/warm as 290 in 2010?	9	N	8	N
4. Which other station(s) has (have) the similar temperature as 290?	9	N	8	N
5. Which year(s) has (have) the similar temperature as 2010?	9	Y,1996	9	Y, 1996
6. Where is/are the coldest place(s) in 2010?	9	N	5	Y, 286
7. When is/are the coldest year(s) in 290?	9	N	9	Y, years in blue cluster
8. Which year(s) is/are the hottest throughout the period in the Netherlands?	9	Y, years in the red cluster	9	Y, years in red cluster
9. What is the pattern of change of temperature in 290 throughout the whole study period?	9	N	9	N
10. Where is the hottest place in the Southeast of the Netherlands from 2005-2008?	9	N	9	N
11. Which year(s) is(are) the hottest between 2000 and 2005 in the northeast of the Netherlands?	8	N	9	N
12. Where is the hottest place between 2000 and 2005 in the whole Netherlands?	9	N	5	Y, 310 in the red cluster
13. What is the hottest temperature in the Netherlands during all years?	9	N	9	(single day, no) Y, average, 11.3
14. What is the pattern of change of temperature in the whole Netherlands?	8	N	7	Y,temperature decreases from SW to NE of the Netherlands

Table 6.5 Agreements on correctness of Station-based SOM

Agreements on correctness of Station-based SOM						
	Station based SOM					
Domain questions	**Cluster map**	**Agreed answer**	**Geo-map**	**Agreed answer**	**Trend plot**	**Agreed answer**
1. What is the lowest temperature in 310 in 2010?	9	N	9	N	8	N
2. What is the lowest temperature in 310 throughout the whole study period?	8	N	9	N	9	Y,lowest mean, 8.7
3. Is/Are there any other place(s) in the Netherlands as cold/warm as 310 in 2010?	7	N	9	N	7	Y,stations in red cluster
4. Which other station(s) has (have) the similar temperature as 310?	9	Y,stations in red cluster	9	Y,stations in red cluster	8	Y,stations in red cluster
5. Which year(s) has (have) the similar temperature as 1992?	9	N	9	N	7	N
6. Where is/are the coldest place(s) in 1992?	9	N	9	N	7	Y, cluster 4
7. When is/are the coldest year(s) in 310?	9	N	9	N	5	Y, 1996
8. Which year(s) is/are the coldest throughout the period in the Netherlands?	9	N	9	N	8	Y, 1996
9. What is the pattern of change of temperature in 310 throughout the whole study period?	8	N	8	N	7	Y, Increasing trend
10. Where is the coldest place in the Southeast of the Netherlands from 2005-2008?	9	N	8	N	9	N
11. Which year(s) is(are) the coldest between 2000 and 2005 in the northeast of the Netherlands?	9	N	9	N	7	Y, 2001
12. Where is/are the coldest places between 2000 and 2005 in the whole Netherlands?	9	N	9	N	7	Y, objects in blue cluster
13. What is the coldest temperature in the Netherlands during all years?	9	N	9	N	8	Y, lowest mean, 8
14. What is the pattern of change of temperature in the whole Netherlands?	8	N	8	Y,Temp decreases from SW to NE	8	Y, An increasing trend

The tables of 6.3, 6.4, 6.5 show:
Most of the answers in the three tables were agreed upon by a majority of respondents, while four questions are doubtful (5 agreements each). They are question 1 in co-clustering, questions 6 and 12 in the year-based SOM trend plot, and question 7 in the station-based SOM trend plot. The reasons for these four answers are explained by the notes and think aloud video recordings taken during the individual experiments.

Question 1 in co-clustering: 5 test users agreed that this question cannot be answered by the graphic, and the explanations were the same and reasonable because in a specific place in a specific year, there was only one attribute value, and no other comparable values. The two reasons for the 4 disagreements are: the test users didn't completely understand the question, and just picked the lowest temperature value of that station; or the searched information led to confusion because the value of the target station in the target year was the lowest average annual value compared to other years.

As for Question 6 and Question12 in the year-based SOM trend plot: each had 5 test users agreed that the question could be answered by the graphic; the search strategies of their answer were correct and answers were the same, were also considered as correct. These two questions are of the same type of questions, requiring searching through both temporal and spatial information. The recordings showed that the 4 users who disagreed with the correct answer did not find the spatial information; moreover, searching time range makes question 12 more complex than question 6, since the year-based trend plot grouped the continuous years into different clusters.

Question 7 in the station-based SOM trend plot: 5 respondents agreed that this question could be answered by the trend plot graphic and gave the same answers; their search strategies were correct and the answers were shown on the graphics. This question is actually the same type of question as questions 6 and 12, requiring both time and spatial information. The other 4 test users disagree because they didn't find the time information.

The doubted questions as well as the questions answered with specific responses and with majority agreements all had the correct information searching strategy, and these answers were validated. Thus, the majority agreements and the doubtful agreements were considered to be the correct answers.

6.7.2. Comparison between generic question results and domain question results, and testing hypotheses

This section compares the results between generic questions and domain question. If the answers to the two sets of corresponding questions are consistent, then the answers to the generic questions are confirmed, and the hypotheses (generated in section 5.4) based on the confirmed questions are verified; otherwise the reasons need to be found.

From the last step, the majority and doubtful answers were confirmed to be the correct answers, according to the "Generic questions visual analysis tables" (Tables 5.2, 5.3, 5.4), "compatibility table" (Table 5.6), and "Agreements on correctness table" (Tables 6.3, 6.4, 6.5), the comparison between the correct answers of the domain questions and the answers to the generic questions are shown in the tables below (Tables 6.6, 6.7, 6.8)
("Y" indicates that the answer to the question can be found in the graphic, while "N" indicates that the question cannot be answered by the corresponding graphic):

Table 6.6 Comparison between generic question results and domain question results- co-clustering

Generic question	co-clustering Geographic map		Domain question	co-clustering Geographic map
1	Y			
2	N		1	N
3	Y		2	Y
4	Y		3	Y
5	Y		4	Y
6	Y		5	Y
7	Y		6	Y
8	Y			
9	Y		7	Y
10	Y		8	Y
11	N		9	N
12	N		10	N
13	Y		11	Y
14	Y		12	Y
15	Y		13	Y
16	N		14	Y

Table 6.7 Comparison between generic question results and domain question results- Year-based SOM

Generic question	Year based SOM			Domain question	Year based SOM	
	Cluster map	Trend plot			Cluster map	Trend plot
1	N	Y				
2	N	N		1	N	N
3	N	Y		2	N	Y
4	N	N		3	N	N
5	N	N		4	N	N
6	Y	Y		5	Y	Y
7	N	Y		6	N	Y
8	N	Y				
9	N	Y		7	N	Y
10	Y	Y		8	Y	Y
11	N	N		9	N	N
12	N	N		10	N	N
13	N	N		11	N	N
14	N	Y		12	N	Y
15	N	Y		13	N	Y
16	N	N		14	N	Y

Table 6.8 Comparison between generic question results and domain question results- Station-based SOM

Generic question	Station based SOM				Domain question	Station based SOM		
	Cluster map	Geo map	Trend plot			Cluster map	Geo map	Trend plot
1	N	N	Y					
2	N	N	N		1	N	N	N
3	N	N	Y		2	N	N	Y
4	N	N	Y		3	N	N	Y
5	Y	Y	Y		4	Y	Y	Y
6	N	N	N		5	N	N	N
7	N	N	Y		6	N	N	Y
8	Y	Y	Y					
9	N	N	Y		7	N	N	Y
10	N	N	Y		8	N	N	Y
11	N	N	Y		9	N	N	Y
12	N	N	N		10	N	N	N
13	N	N	Y		11	N	N	Y
14	N	N	Y		12	N	N	Y
15	N	N	Y		13	N	N	Y
16	N	N	Y		14	N	Y	Y

From the tables of "comparison between generic question results and domain question results" (Tables 6.6, 6.7, 6.8):

In co-clustering, all answers to the generic questions can be confirmed by the answers to the corresponding domain questions except questions 1, 8 and 16. The answer to question 1 is very obvious, being about the attribute value of a given year in a given place. To answer question 8, respondents only have to search for the highest attribute value in the visual outputs. As for question 16 is that the researcher didn't notice that the geo-information can reflect trend information.

In year-based SOM, still the questions 1, 8 and 16 still weren't confirmed by the answers to the domain questions. For question 1, the cluster map doesn't offer spatial information, and the trend plot offers both time and spatial information. The answer is quite obvious from the graphics. For question 8, the respondent needs to find the highest point (station with the highest temperature) in year clusters; the cluster map does not contain spatial information, and the trend plot can obviously show the highest point. And for question 16, it is because the way the information is provided with the station points on the lines arranged from the southwest to the northeast of the Netherlands is ignored. The answer should be the same as the domain questions.

In the station-based SOM, the unconfirmed questions are still the question 1, 8 and 16. For question 1, the cluster map and geo-map don't have time information, and the trend plot shows the required information clearly. For question 8, one only needs to find the station cluster with the highest attribute value. This information can easily be seen on all three graphics of the station-based SOM. The reason for question 16 is the same as the co-clustering geo-map; the geo-information was ignored. The answer should be the same as the domain questions.

The comparison result only changed the hypotheses generated from generic question 16.
Based on the above analysis and interpretation, the hypotheses (generated in section 5.4) are verified and formulated into three tables below. In the tables, the time, location and attribute information were each separated into subcategories, according to the different "search range" and "search target" appearing in the questions. In this way, one can see if a question can be answered (Y/N) in a given search range, and if the search range can be found (1/0). ("1" and "0" refer to search range, if the search range can be found, the item is marked as "1", otherwise "0"; "Y" and "N" refer to the search target, if the needed information can be answered, it is marked as "Y", otherwise mark it as "N")

Table 6.9 Suitability in answering questions- co-clustering

Co-clustering	Generic question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Attribute value	Single value		N														
	Mean value	Y										1					
	Highest /lowest mean value			Y				1	1	1	1		0	1	1	Y	
Time	Single year	1	1		1			1									
	Similar years						Y			Y	Y			Y			
	Time period												1		1		
	Whole period								1	1	1					1	1
Spatial	Single place	1	1	1						1		1	N				
	Similar places				Y	Y		Y	Y						Y		
	Geographic area												1	1			
	Whole area				1			1	1							1	1
Trend and pattern	Time perspective											N					N
	Spatial perspective											N					Y

The table 6.9 shows the verified hypotheses of the co-clustering visual outputs, describing the suitability of the tested co-clustering visual outputs in answering questions.
Within the different searching ranges that are embedded in spatial, temporal, and attribute aspects:
The search targets that were answered ("Y"s) are: similar place, similar years, mean value and highest mean value, and general trend in spatial perspective.
The search targets that were not answered ("N"s) are: single value, single place from an area, and general trend in time perspective, trend of a single place in time or spatial perspective.

The answers focus on a year or place cluster, or the attribute value of the year or place cluster.
No trend information about a single place was discovered; but the spatial trend of the whole area over years was found. The visual outputs didn't show a location or year difference inside a location or year cluster. The original single value cannot be found.

Table 6.10 Suitability in answering questions- Year-based SOM

Year-based SOM	Generic question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Attribute value	Single value		N														
	Mean value	Y										1					
	Highest /lowest mean value			Y				1	1	1	1		N	1	1	Y	
Time	Single year	1	1		1			1						N			
	Similar years						Y			Y	Y						
	Time period												1		1		
	Whole period								1	1	1	1				1	1
Spatial	Single place	1	1	1				Y	Y	1		1			Y		
	Similar places				N	N											
	Geographic area												0	0			
	Whole area				1			1	1							1	1
Trend and pattern	Time perspective											N					N
	Spatial perspective											N					Y

The table 6.10 shows the verified hypotheses of the year-based SOM visual outputs, describing the suitability of the tested year-based SOM visual outputs in answering questions.
Within the different searching ranges that are embedded in spatial (exclude geo-area), temporal, and attribute aspects:
The search targets that were answered ("Y"s) are: similar years, single place, mean value and highest mean value, and general trend in spatial perspective.
The search targets that were not answered ("N"s) are: similar place, single original value, and general trend in time perspective, and trend of a single place in time or spatial perspective.

The answers focus on a year cluster or a single place, or the attribute value of the year cluster or single place. No trend information was found about a single place over years; but the spatial trend of the whole area over years was found. The visual outputs didn't show years difference inside a year cluster. The original single value cannot be found.

Table 6.11 Suitability in answering questions- Station-based SOM

Station-based som	Generic question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Attribute value	Single value		N														
	Mean value	Y										1					
	Highest /lowest mean value			Y				1	1	1	1		N	1	1	Y	
Time	Single year	1	1		1			1		Y	Y			Y			
	Similar years						N										
	Time period												1		1		
	Whole period								1	1	1	1				1	1
Spatial	Single place	1	1	1						1		1					
	Similar places				Y	Y		Y	Y						Y		
	Geographic area												1	1			
	Whole area				1			1	1							1	1
Trend and pattern	Time											Y					Y
	Spatial											N					Y

The table 6.11shows the verified hypotheses of the station-based SOM visual outputs, describing the suitability of the tested station-based SOM visual outputs in answering questions.
Within the different searching ranges that are embedded in spatial, temporal, and attribute aspects:
The search targets that were answered ("Y"s) are: similar place, single year, mean value and highest mean value, trend of whole study area, trend of whole time period, trend of a single place in time perspective.
The search targets that were not answered ("N"s) are: similar years, single original value, and trend of a single place in spatial perspective.

The answers focus on a spatial cluster or a single year, or the attribute value of the spatial cluster or the single year.
The trend information about the in temporal aspect of a single place over years can be found; the spatial and temporal trends of the whole area over years were found. The visual outputs didn't show locations difference inside a station cluster. The original single value cannot be found.

6.7.3. Satisfaction

The information about satisfaction of the tested visual graphics was collected from the Likert scale questionnaires from nine test users. The statistics result is shown in the table below (Table 6.12):

Table 6.12 Likert scale questionnaire statistics

Likert scale questionnaire statistics the titles of S-D, D, M, A, S-A, Val, and Perc, indicate Strongly Disagree, Disagree, Moderate, Agree, Strongly Agree, Value, and Percentage respectively.

Likert scale questionnaire statistics									
Aspect	**Graphic**	**Frequency**	**S-D**	**D**	**M**	**A**	**S-A**	**Val**	**Perc**
	Cluster map	Q1 Do you like the cluster maps of year-based and station-based SOM?		1	3	2	3	34	0.76
	Trend plot	Q2 Do you like the trend plots of year-based and station-based SOM?		3	3		3	30	0.67
Pleasant to see	Geo-map	Q3 Do you like the geo-maps of station-based SOM and co-clustering?		1	1	2	5	38	0.84
	Cluster map	Q4 Do you think the cluster maps of year-based and station-based SOM are designed properly?	1	2	2	2	2	29	0.64
	Trend plot	Q5 Do you think the trend plots of year-based and station-based SOM are designed properly?	1	1	5	2		26	0.58
Proper design	Geo-map	Q6 Do you think the geo-maps of station-based SOM and co-clustering are designed properly?			1	3	5	40	0.89
	Cluster map	Q7 Do you think the cluster maps of year-based and station-based SOM are easy to understand?	1		2	2	4	35	0.78
	Trend plot	Q8 Do you think the trend plots of year-based and station-based SOM are easy to understand?		4	2	1	2	28	0.62
Ease of use	Geo-map	Q9 Do you think the geo-maps of station-based SOM and co-clustering are easy to understand?			2	3	4	38	0.84

In the table 6.12, there are five levels on the Likert scale for each question, and 9 marks for each question, from 9 questionnaires. The value of the lowest level is 1; of the highest level is 5; the intervals between continuous levels are the same. The maximum value for each question is 45, if all the test persons strongly

agree with the statement (9*5). The frequency of each value for each question was counted, and the values obtained for each question were calculated, as well as accounting for the possible maximum value.
The percentage indicates the preference level. The higher level represents a higher preference, or agreement.
Based on this table, the histogram below was generated, from which the results from this table can be clearly seen, showing the three aspects of "pleasant to see," "proper design," and "ease of use."

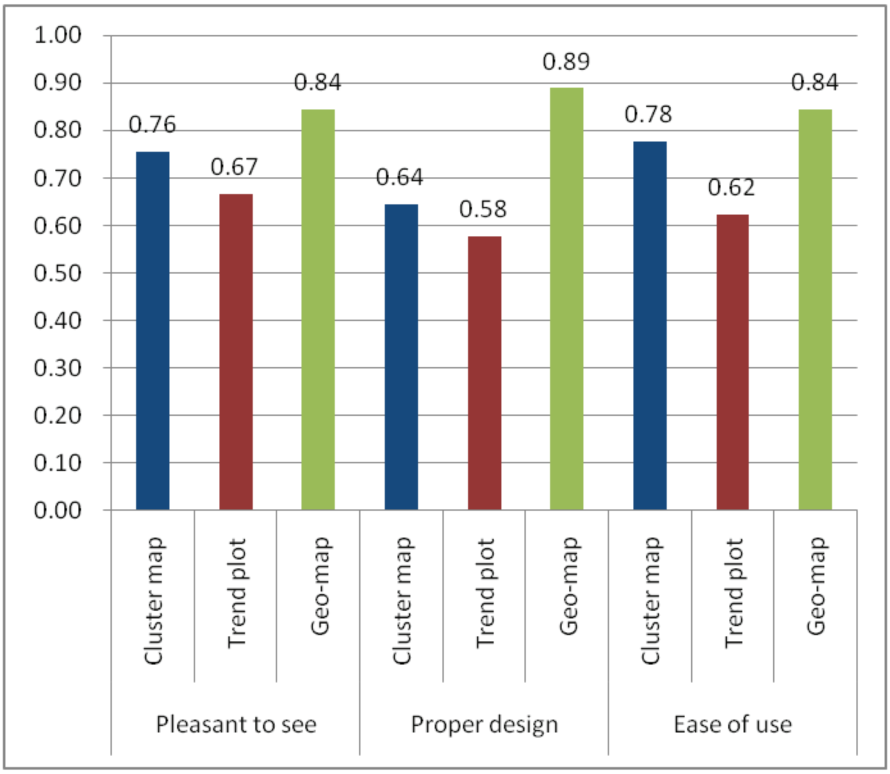


Figure 6.1 Histogram of Likert scale questionnaire statistics

The figure 6.1 shows the geographic maps from station-based SOM and co-clustering all scored the highest percentages in the three aspects, 84%, 89% and 84%, respectively. The reasons of test users made these choices on likert scale questionnaire were collected from the focus group, explaining that the test users think the geo-map is pleasant to see and the colours are used properly; the design reflected spatial and temporal information clearly, and the way the information is shown to the users is very straightforward and easy to understand. The drawback is that the legends need improvement.

The cluster maps from the year-based SOM and the station-based SOM are in the middle level of the three types of graphics. Test users think the colours of the cluster map are used properly. There were confusions because the sizes of clusters are different, and the objects within each cluster are not distributed evenly. Another problem is that the names of the clusters on the scale bar are not ordered. The cluster maps were considered easy to use but contained too little information to answer questions.

The trend plot from the year-based SOM and the station-based SOM are at the lowest level because they were considered difficult to understand; the legend of the diagrams should have more details in order to reduce the amount of time taken and remove some misunderstandings. The advantage to this design is that it includes both spatial and temporal information; and serves as a supplement to the year cluster map by offering spatial information, and to the station cluster map by offering temporal information.

6.7.4. Effectiveness

The effectiveness in this context was measured in two levels: whether the questions could be answered, and how well the visual outputs answer the domain questions. "1" indicates the question can be answered by the corresponding graphic, while "0" indicates the opposite.

Table 6.13 Questions answered by the three types of visual outputs

Question answered by the three ways of visual outputs						
	co-clustering	**Year based SOM**		**Station based SOM**		
Domain questions	**Geo-map**	**Cluster map**	**Trend plot**	**Cluster map**	**Geo- map**	**Trend plot**
1	0	0	0	0	0	0
2	1	0	1	0	0	1
3	1	0	0	0	0	1
4	1	0	0	1	1	1
5	1	1	1	0	0	0
6	1	0	1	0	0	1
7	1	0	1	0	0	1
8	1	1	1	0	0	1
9	0	0	0	0	0	1
10	0	0	0	0	0	0
11	1	0	0	0	0	1
12	1	0	1	0	0	1
13	1	0	1	0	0	1
14	1	0	1	0	1	1
Total	11	8		11		
Percentage	0.79	0.57		0.79		

The table 6.13 shows which question could be and were answered by each visual output.

The table 6.14 below calculated the number of respondents that correctly answered the questions that could be answered.

Table 6.14 Effectiveness

Agreements on correctness- Effectiveness						
	Co-clustering	Year-based SOM		Station-based SOM		
Domain questions	Geo-map	Cluster map	Trend plot	Cluster map	Geo-map	Trend plot
1	5	9	8	9	9	8
2	8	9	9	8	9	9
3	9	9	8	7	9	7
4	9	9	8	9	9	8
5	9	9	9	9	9	7
6	8	9	5	9	9	7
7	9	9	9	9	9	5
8	9	9	9	9	9	8
9	9	9	9	8	8	7
10	7	9	9	9	8	9
11	8	8	9	9	9	7
12	8	9	5	9	9	7
13	9	9	9	9	9	8
14	9	8	7	8	8	8
Total red value	95	18	62	9	17	81
Effectiveness	0.96	1	0.86	1	0.94	0.82

Table 6.14 shows the number of correct answers of each graphic with each question. Based on table 6.13, the intersection numbers were coloured red; they are the correct answers in all the questions that can be answered. The total correct answers for each graphic were calculated and then calculated their possible maximum percentage- effectiveness.

In the three types of visual outputs, each of the six graphics has different effectiveness:

The cluster maps in the year-based SOM and in the station-based SOM have the highest effectiveness. The feedback from the users explained: the cluster maps are easy to use, and contained little information to answer questions.

The co-clustering geographic map was the second highest, 0.96. The feedback about this graphic was the way it represent the information was very straightforward.

The trend plots in the station-based SOM and year-based SOM scored the lowest and the second lowest number, 0.86 and 0.82 respectively. Again from the users' feedback, they think this kind of graphic are difficult to understand,

It is difficult to compare the effectiveness between visual outputs groups, but if from the average perspective, co-clustering would score the highest effectiveness.

6.7.5. Efficiency

The efficiency here is the time required to answer questions correctly, it is measured in three aspects: the questions answered according to the three types of visual outputs, the time cost and the effectiveness.
According to the answered domain question forms, the table 6.13 shows the question could be and were answered correctly by each visual output.
"1" indicates the answer has been answered by the corresponding graphic, "0" indicates the opposite.
The questions answered by each type of clustering visual outputs were calculated.
Out of 14 domain questions, the questions that have been answered by the three ways of clustering visual outputs are: 11 questions for the combination of visual output of the station-based SOM, 8 questions for the combination of visual outputs of the year-based SOM and 11 questions for the combination of visual output of the co-clustering geographic map, which account for 79%, 57% and 79% respectively.

The time cost table 6.15 is based on the records of the time costs of each domain question referring to different graphics, involving each test user.

Table 6.15 Time taken on individual experiment

Time cost table (unit: second)						
Graphic	Co-clustering	Year based SOM		Station based SOM		
Domain Question	Geo-map	Cluster map	Trend plot	Cluster map	Geo-map	Trend plot
Q1	42.6	10.9	31.9	12.8	14.8	31.8
Q2	24.0	10.1	20.0	11.0	10.0	24.6
Q3	21.4	8.6	37.6	16.9	15.0	25.2
Q4	24.3	7.1	29.9	12.1	14.9	9.9
Q5	22.7	8.2	26.9	8.4	6.6	56.6
Q6	15.2	6.3	19.6	7.2	4.6	17.4
Q7	22.8	7.8	29.7	6.7	5.0	21.1
Q8	10.3	11.7	19.2	8.1	9.6	25.2
Q9	32.2	12.2	32.2	14.1	9.2	35.0
Q10	28.3	8.7	25.9	15.2	14.1	27.2
Q11	29.0	12.4	41.1	10.6	6.8	45.1
Q12	29.8	8.0	45.3	10.1	9.9	18.0
Q13	18.9	10.9	17.9	9.2	11.9	17.8
Q14	25.7	15.7	20.0	14.1	16.1	16.4
Average total time per graphic	347.2	138.6	397.1	156.6	148.3	371.3
Average total time per visual outputs	347.2	535.7		676.2		

In the table 6.15, the average time consumed by nine test persons on each graphic of all the domain questions is calculated. The calculation shows that the time costs for the trend plot from Year-based SOM, the trend plot from station-based SOM, and the geographic map from co-clustering are the most, more than twice the rest of the graphics. This result can be explained by the Likert scale analysis and the table of "Questions answered by the three ways of visual outputs," which explains that these three graphics are more complex, containing more information than the other ones. In these three ways of visual outputs, co-clustering visual outputs took least time and station-based SOM cost the most. The reason was inferred from the Likert scale questionnaire and the focus group that the co-clustering geographic map represents information in a straightforward way.

Based on the above analysis, the efficiency table (Table 6.16) is created.

Table 6.16 Efficiency

Efficiency						
Graphic	Co-clustering	Year-based SOM		Station-based SOM		
Aspect	Geo-map	cluster map	Trend plot	cluster map	geo-map	Trend plot
Questions have been answered	79%	57%		79%		
Time cost	347	535.7		676.2		

It shows that the co-clustering visual outputs answered the most domain questions and took the least time. Moreover, referring to the result from effectiveness step that co-clustering visual outputs scored the highest effectiveness amongst of these three types of visual outputs, the conclusion about the efficiency was made that co-clustering visual outputs scored the highest efficiency.

Detail analysis of time table

When designing the user test experiment, it was considered that users may learn from analyzing the first group of visual outputs, and then they might perform faster for the second or the third group of visual outputs. And that would bring the bias when register the time taken to different groups. So, in individual experiments, the domain question forms were designed in three different orders according to the three test groups to reduce the order effects.

From the time records of each group, this influence can be seen in the tables (tables 6.17, 6.18, and 6.19) below.

Table 6.17 Time records of individual experiment- the first order

Time records- the first order						
Graphic	**Co-clustering**	**Year based SOM**		**Station based SOM**		
Domain question	**Geo-map**	**Cluster map**	**Trend plot**	**Cluster map**	**Geo-map**	**Trend plot**
Q1	47.0	17.0	28.7	9.7	26.7	21.3
Q2	27.3	17.3	17.0	10.7	10.0	24.0
Q3	33.7	12.3	16.7	15.0	12.3	29.0
Q4	40.3	11.3	9.0	15.7	14.7	12.3
Q5	28.3	10.0	46.3	10.7	7.0	72.0
Q6	14.7	9.0	13.0	12.3	5.7	13.0
Q7	18.7	11.7	43.3	7.0	5.7	16.0
Q8	13.3	10.7	30.0	8.7	8.3	26.0
Q9	39.7	18.3	10.7	10.0	9.3	48.7
Q10	19.0	11.0	15.3	13.0	9.3	19.7
Q11	25.7	18.0	8.0	16.0	6.7	41.7
Q12	21.7	9.0	17.0	14.3	9.3	18.7
Q13	23.7	17.3	13.3	12.0	12.3	11.3
Q14	20.7	7.7	19.7	6.7	10.7	15.7
Total	373.7	180.7	288.0	161.7	148.0	369.3

This table is the first order for the first group. Test persons answered the questions by column from left to right.

Table 6.18 Time records of individual experiment- the second order

Time records- the second order						
Graphic	**Year based SOM**		**Station based SOM**			**Co-clustering**
Domain question	**Cluster map**	**Trend plot**	**Cluster map**	**Geo-map**	**Trend plot**	**Geo- map**
Q1	5.7	32.3	14.3	4.3	31.3	37
Q2	7.3	20.3	6.0	4.7	14.3	15.3
Q3	7.7	62.0	21.3	17.7	33.0	19.3
Q4	4.0	70.7	7.3	6.3	10.7	18.3
Q5	5.3	12.0	8.0	6.7	33.0	13.7
Q6	4.7	14.3	4.0	3.7	15.0	11.7
Q7	5.3	20.0	3.7	3.7	18.3	17.0
Q8	13.0	15.3	7.7	8.7	27.0	7.7
Q9	6.3	28.0	7.3	5.7	21.3	24.0
Q10	5.0	29.0	23.0	19.3	40.0	31.7
Q11	9.0	61.0	6.0	5.0	37.0	27.0
Q12	5.0	34.0	7.7	11.0	15.0	31.7
Q13	5.0	22.7	4.0	7.3	25.0	12.0
Q14	5.3	23.0	4.7	12.3	17.7	26.3
Total	88.7	444.7	125.0	116.3	338.7	292.7

This table is the second order for the second group. Test persons answered the questions by column from left to right.

Table 6.19 Time records of individual experiment-the third order

Time records- the third order						
Graphic	**Station based SOM**			**Co-clustering**	**Year based SOM**	
Domain question	**Cluster map**	**Geo-map**	**Trend plot**	**Geo-map**	**Cluster map**	**Trend plot**
Q1	14.3	13.3	42.7	43.7	10	34.7
Q2	16.3	15.3	35.3	29.3	5.7	22.7
Q3	14.3	15.0	13.7	11.3	5.7	34.0
Q4	13.3	23.7	6.7	14.3	6.0	10.0
Q5	6.7	6.0	64.7	26.0	9.3	22.3
Q6	5.3	4.3	24.3	19.3	5.3	31.3
Q7	9.3	5.7	29.0	32.7	6.3	25.7
Q8	8.0	11.7	22.7	10.0	11.3	12.3
Q9	25.0	12.7	35.0	33.0	12.0	58.0
Q10	9.7	13.7	22.0	34.3	10.0	33.3
Q11	9.7	8.7	56.7	34.3	10.3	54.3
Q12	8.3	9.3	20.3	36.0	10.0	85.0
Q13	11.7	16.0	17.0	21.0	10.3	17.7
Q14	31.0	25.3	16.0	30.0	34	17.3
Total	183.0	180.7	406.0	375.3	146.3	458.7

This table is the third order for the third group. Test persons answered the questions by column from left to right.

The above three tables show an influence of the order of presentation of the outputs.
When the geographic map of co-clustering was presented first, test persons used far more time than when it was presented last (373.1compare to292.7).
When the cluster map of station based SOM was presented from the first to the last, the time taken was: 183, 125, 161.
When the trend plot of station based SOM was presented from the first to the last, the time taken was: 406, 338.7, 369.3.
It was found that in most cases, the graphics were presented first took longer time than they were presented later. But this was not for the entire situation. For example, cluster map in year based SOM took 88 seconds, 180 seconds, and 146.3 seconds, when it was present from the first to last. Considering the analysis of test users' information (table 6.20 below), the reason were found that the second group was the fastest group.

6.7.6. Test user information analysis

The user information might also influence the test result.

Table 6.20 User information analysis

Test participants analysis				
Participant	**Present situation**	**Correct answers**	**percentage**	**Time cost (unit: minute)**
p1	M.Sc in GIP	80	0.95	35
p2	M.Sc in UPM	79	0.94	22.6
p3	M.Sc in NRM	75	0.89	18.5
p4	M.Sc in GIP	81	0.96	33.1
p5	PhD, ITC Earth Systems' Analysis	80	0.95	22.1
p6	M.Sc in GIP	81	0.96	15.1
p7	M.Sc in NRM	72	0.86	22.8
p8	M.Sc in WRM	78	0.93	28.8
p9	PHD in GIP	76	0.90	36.0

The answers to the domain questions from each user were analyzed. The answers regarded as correct were counted. The possible maximum percentages were calculated for the convenience of comparison, ranging from 86% to 96%.

Comparing these results to users' experience, we found the better results are from the test users who have geo-informatics and cartographic backgrounds (p1 95%, p4 96%, and p6 96%), and professional experiences in analyzing spatiotemporal data (p5 95%), P5 and P6 took the least time. This is considered a fact that work experience influences the analysis result. However, the percentages of correctness of the results from other test users from other departments are also good. This means that the selected clustering visual outputs can be used to derive information from big spatiotemporal data by researchers without professional experience.
The two fasted persons (P5 and P6) were both in the group two, this information explained why the group 2 took the least time and made the order influence didn't look obvious.

6.7.7. Conclusion

The experiment was implemented in this chapter. From the implementation result, the hypotheses were tested, from which the visual outputs of year-based SOM, station-based SOM and co-clustering are able to answer what kind of questions were analyzed, and the results are formulated into three tables; the satisfaction, effectiveness and efficiency of the visual outputs of the year-based SOM, station-based SOM and co-clustering were compared and discussed; based on the comparison and discussion, the drawback of the selected visual outputs can be improved; and the test participants' information was analyzed to reveal the influence factor of the experiment and offer potential users an impression when doing this kind of spatiotemporal data analysis.

7. FINAL CONCLUSION AND RECOMMENDATION

7.1. Conclusion

This research described the suitability of two clustering methods to answer questions related to spatiotemporal big data by using their visual outputs, based on the case study spatiotemporal weather dataset. This was achieved by accomplishing four goals: generating generic spatiotemporal questions, establishing the characteristics of the visual outputs from the two clustering methods, finding the domain questions of the case study data from analysis experts, and evaluating the strengths and weaknesses of the chosen clustering visual outputs.

Generic questions were generated based on the nature and inherent properties of spatiotemporal big dataset, and based on a scheme to make sure the generated generic questions cover different information search levels. The similar questions that differed in the intermediate and overall search levels were combined together, and the questions were organized in increasing order based on the complexity of the analysis task. After defining the generic questions, a spatiotemporal big dataset (the Netherlands 20-year daily temperature data) was selected as the case study data; and the typical visual outputs that represent the characteristics of the clustering technique and the case study dataset were selected from the visual outputs of SOM(year-based SOM and station-based SOM) and co-clustering, respectively. These selected visual outputs answered the generic questions differently; based on the answers, the hypotheses about the suitability of the selected clustering visual outputs answering certain kinds of spatiotemporal generic questions were formulated. And then the domain questions were obtained from the domain experts, based on the case study dataset. These domain questions were combined and organized in the same way as the generic questions, and verified to have consistency with the generic questions so that the answers to the domain questions could be used in the experiment for testing the hypotheses.

An information-seeking task (users search for information in the selected visual outputs to answer domain questions), and think aloud (users speak their thoughts about the test) were involved in the user experiment.

when answers to the corresponding domain question were agreed upon by the majority of test users, and after analyzing and interpreting were regarded as being valid, then the answer was regarded as correct; when answers to the same question were agreed upon by a doubtful number of test users, analysis and interpretation were applied to find the reason, and then to determine if it was correct. The correct answers were compared to the answers of the generic questions, and then to determine the hypotheses were tested. The user research methods of observation and recording (use video and notes to record the process objectively, and registered the time for answering each question), questionnaire (collect opinions from test subjects about selected visual outputs) and focus group (explain about the answers on the questionnaires) were used to collect the experiment data and evaluate the strengths and weaknesses of the selected visual outputs in answering domain questions.

7.2. Recommendation

The selected three ways of cluster visual outputs can all offer spatial and temporal information from spatiotemporal big dataset. They have two points in common: questions cannot be answered if they are about searching for the attribute values of original data, or if they are about finding the difference between the objects inside a year or station cluster. These are decided by the nature of the clustering technique.
Potential users who want to choose a clustering technique based on their visual outputs should depend on their differences in answering questions. The comparison is shown in the table below; it is based on users'

feedback for the experiment, and analysis and interpretation of the data(answers to the domain questions, video recording, notes taken in individual experiment, Likert questionnaire, notes from the focus group) collected from the experiment.

Table 7.1 The three types of visual outputs comparison

Visual outputs comparison- differences			
		SOM	
Method Aspect	**co-clustering**	**year-based**	**station based**
Graphic	Geographic map	Cluster map, trend plot	Cluster map, trend plot, geographic map
User preference	More	Less	
The way it organizes information	Organize time and location information simultaneously, take much less time	Organize time and location information separately, take more time	
Easy to understand	Easier	Difficult	
Effectiveness	Higher	Lower	
Offer similar information	Both time and location	Only time	Only location
The answers	Are about a year or station cluster, or the attribute value of the year or the station cluster.	About a single year or a station cluster, or the attribute of a single year or a station cluster	About a single station or a year cluster, or the attribute value of a single station or year cluster
Offer geo-information	Yes	No	Yes
Trend/pattern	Geographic perspective	Geographic perspective	Both Geographic and time perspective

Overall, both of the SOM and co-clustering techniques can be used to this kind of spatiotemporal temperature big dataset to obtain spatial and temporal information. In cases of different requirements, based on the comparison of the selected visual outputs, the recommendations are:

1. If users want to see similarity of time and location simultaneously, co-clustering is suitable.
2. If users want to know the similarity of time and location quickly, co-clustering is suitable.
3. If users prefer to use geographic map, co-clustering is suitable.
4. If users want to see differences between stations inside a year cluster, year-based SOM is suitable.

5. If users want to see differences between years inside a station cluster, station-based SOM is suitable.
6. If users want to know the change over time, station-based SOM is suitable.
7. If users want to see the geographic trend/pattern, both methods are suitable.

LIST OF REFERENCES

Agarwal, P., & Skuping, A. (2008). *Self organising maps : applications in geographic information science*. Chichester: Wiley & Sons.

Andrienko, G, Andrienko, N., Bremm, S., Schreck, T., Von Landesberger, T., Bak, P., & Keim, D. (2010). Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Computer Graphics Forum, 29*(3), 913-922.

Andrienko, Natalia, & Andrienko, Gennady. (2013). A visual analytics framework for spatio-temporal analysis and modelling. *Data Mining and Knowledge Discovery, 27*(1), 55-83.

Andrienko, Natalia, Andrienko, Gennady, & Gatalsky, Peter. (2003). Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing, 14*(6), 503-541. doi: http://dx.doi.org/10.1016/S1045-926X(03)00046-6

Bertin, Jacques. (1983). *Semiology of graphics : diagrams, networks, maps*. Madison, Wis.; London: University of Wisconsin Press.

Bock, Hans-Hermann. (2011). Special Issue on 'Time series clustering'. *Advances in Data Analysis and Classification, 5*(4), 247-249.

Clayton Lewis, John Rieman. (1993). Task-Centered User Interface Design Vol. Testing The Design With Users. *A Practical Introduction* Retrieved from http://hcibib.org/tcuid/index.html#notice

Deng, Min, Liu, QiLiang, Wang, JiaQiu, & Shi, Yan. (2011). A general method of spatio-temporal clustering analysis. *Science China Information Sciences*, 1-14.

Dhillon, Inderjit S., Mallela, Subramanyam, & Modha, Dharmendra S. (2003). *Information-theoretic co-clustering*. Paper presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C.

Dobre, C., & Xhafa, F. Intelligent services for Big Data science. *Future Generation Computer Systems*(0). doi: http://dx.doi.org/10.1016/j.future.2013.07.014

Estivill-Castro, Vladimir. (2002). Why so many clustering algorithms: a position paper. *SIGKDD Explor. Newsl., 4*(1), 65-75.

Focus groups. Retrieved from usabilitynet website: http://www.usabilitynet.org/tools/focusgroups.htm

Frøkjær, Erik, Hertzum, Morten, & Hornbæk, Kasper. (2000). *Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?* Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Fu, Tak-chung. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence, 24*(1), 164-181.

Hirano, S., & Tsumoto, S. (2005). Empirical comparison of clustering methods for long time-series databases. In S. Tsumoto, T. Yamaguchi, M. Numao & H. Motoda (Eds.), *Active Mining* (Vol. 3430, pp. 268-286).

Hossain, M. Shahriar, Ramakrishnan, Naren, Davidson, Ian, & Watson, LayneT. (2012). How to "alternatize" a clustering algorithm. *Data Mining and Knowledge Discovery*, 1-32.

Hudson, IreneL, Keatley, MarieR, & Lee, ShalemY. (2011). Using Self-Organising Maps (SOMs) to assess synchronies: an application to historical eucalypt flowering records. *International Journal of Biometeorology, 55*(6), 879-904.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv., 31*(3), 264-323.

Kaplan, Bonnie, & Maxwell, JosephA. (2005). Qualitative Research Methods for Evaluating Computer Information Systems. In J. Anderson & C. Aydin (Eds.), *Evaluating the Organizational Impact of Healthcare Information Systems* (pp. 30-55): Springer New York.

Keim, Daniel, Andrienko, Gennady, Fekete, Jean-Daniel, Görg, Carsten, Kohlhammer, Jörn, & Melançon, Guy. (2008). Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. Stasko, J.-D. Fekete & C. North (Eds.), *Information Visualization* (Vol. 4950, pp. 154-175): Springer Berlin Heidelberg.

Kleinberg, Jon. (2003). An impossibility theorem for clustering. *Advances in neural information processing systems*, 463-470.

Koua, Etien Luc. (2005). Computational and visual support for exploratory geovisualization and knowledge construction. *ITC Dissertation; 118*.

Lai, H. P., Visani, M., Boucher, A., & Ogier, J. M. (2012). An experimental comparison of clustering methods for content-based indexing of large image databases. *Pattern Analysis and Applications, 15*(4), 345-366.

Lu, Yonggang, & Wan, Yi. (2012). Clustering by Sorting Potential Values (CSPV): A novel potential-based clustering method. *Pattern Recognition, 45*(9), 3512-3522.

Marinai, Simone, Marino, Emanuele, & Soda, Giovanni. (2008). *A Comparison of Clustering Methods for Word Image Indexing.* Paper presented at the Document Analysis Systems.

Meil, Marina. (2005). *Comparing clusterings: an axiomatic view.* Paper presented at the Proceedings of the 22nd international conference on Machine learning, Bonn, Germany.

Meila, M., & Heckerman, D. (2001). An experimental comparison of model-based clustering methods. *Machine Learning, 42*(1-2), 9-29. doi: 10.1023/a:1007648401407

Mirkin, B.G. (1996). *Mathematical Classification and Clustering*: Kluwer Academic Publishers.

Moghaddam, Saeed, Helmy, Ahmed, Ranka, Sanjay, & Somaiya, Manas. (2010). *Data-driven co-clustering model of internet usage in large mobile societies.* Paper presented at the Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems.

Mount, N. (2009). Self-organising Maps: Applications in Geographic Information Science. [Book Review]. *International Journal of Geographical Information Science, 23*(4), 545-546. doi: 10.1080/13658810802344168

Nielsen, J. (1997). The use and misuse of focus groups. *Software, IEEE, 14*(1), 94-95. doi: 10.1109/52.566434

Nielsen, Jakob. (1994). *Usability inspection methods.* Paper presented at the Conference Companion on Human Factors in Computing Systems, Boston, Massachusetts, USA.

NIELSEN, JAKOB. (2000). Why You Only Need to Test with 5 Users. Retrieved from Nielsen Norman Group website: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

NIELSEN, JAKOB. (2003). Recruiting Test Participants for Usability Studies. Retrieved from Nielsen Norman Group website: http://www.nngroup.com/articles/recruiting-test-participants-for-usability-studies/

NIELSEN, JAKOB. (2012). How Many Test Users in a Usability Study? Retrieved from Nielsen Norman Group website: http://www.nngroup.com/articles/how-many-test-users/

Peuquet, Donna 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

Razeghi, Rozita. (2010). Usability of Eye Tracking as a User Research Technique in Geo-information Processing and Dissemination.

Rokach, Lior, & Maimon, Oded. (2005). Clustering Methods. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 321-352): Springer US.

Sagiroglu, S., & Sinanc, D. (2013, 20-24 May 2013). *Big data: A review.* Paper presented at the Collaboration Technologies and Systems (CTS), 2013 International Conference on.

Shneiderman, B. (1996). *The eyes have it: A task by data type taxonomy for information visualizations*. Los Alamitos: I E E E, Computer Soc Press.

. Usability Design and Evaluation. (2006) *Designing Human Interface in Speech Technology* (pp. 123-165): Springer US.

van Elzakker, C.P.J.M., Ormeling, F.J., & Kraak, M.J. (2004). *The use of maps in the exploration of geographic data.* (116), University of Utrecht, Utrecht. Retrieved from http://www.itc.nl/library/Papers_2004/phd/vanelzakker.pdf

Van Mechelen, I., Bock, H. H., & De Boeck, P. (2004). Two-mode clustering methods: a structured overview. [Article]. *Statistical Methods in Medical Research, 13*(5), 363-394. doi: 10.1191/0962280204sm373ra

Wang, Haiying, Zheng, Huiru, & Azuaje, Francisco. (2008). Clustering-based approaches to SAGE data mining. *BioData Mining, 1*(1), 1-12.

Wang, Xiaozhe, Smith, Kate, & Hyndman, Rob. (2006). Characteristic-Based Clustering for Time Series Data. *Data Mining and Knowledge Discovery, 13*(3), 335-364.

Wei, Xu, & Hou-kuan, Huang. (2006, Aug. 30 2006-Sept. 1 2006). *Research and Application of Spatio-temporal Data Mining Based on Ontology.* Paper presented at the Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on.

Wilkinson, Leland, & Friendly, Michael. (2009). The history of the cluster heat map. *The American Statistician, 63*(2).

Wu, Meng-Lun, Chang, Chia-Hui, & Liu, Rui-Zhe. (2012). Co-clustering with augmented matrix. *Applied Intelligence*, 1-12.

Wu, Sitao, & Chow, TommyW S. (2003). Self-Organizing-Map Based Clustering Using a Local Clustering Validity Index. *Neural Processing Letters, 17*(3), 253-271.

APPENDIX 1

Interview email

Dear data processing experts,

As we communicated before, for my MSC research, I need to get domain questions (real information requirements) for my case study dataset. It is Netherlands 20 years (from 1992 to 2011) daily temperature data from twenty-eight Dutch meteorological stations. The station IDs are: 310, 319, 323, 330, 344, 210, 350, 348, 240, 380, 370, 356, 260, 375, 391, 235, 275, 269, 267, 273, 278, 270, 283, 277, 279, 290, 280, and 286.

The question and its requirement are listed below:
1. What information do you want to know from the above context data? (The proposed questions should be related to the real problem)
2. Could you please formulate your questions from the aspects of attribute (what), time (when) and location (where), and from different levels of elementary, intermediate and overall?

(The elementary questions are about local questions, they are specific to the object or time or location point. The intermediate questions are about range, for example, information from a period of time or parts of the study area. The overall questions are about the whole study area and the whole time span.)

Thank you very much for your understanding and support.

Kind regards

Qian

APPENDIX 2

Invite participants email

Dear all,

For my research study, I need several test participants for the experiment phase. As your research is, about spatiotemporal big data which is relevant to this research study, I would like to invite you to be one of the test users.

The experiment is for finding out how the visual outputs of the selected clustering methods answer specified questions from a big spatiotemporal dataset with daily temperature data from the Netherlands for a period of 20 years. The result would be useful, to potential users who want to choose suitable data processing methods, for retrieving a certain type of information from big spatiotemporal datasets.
The test is for examining the visual outputs, to see what kind of information they can offer. So, if you cannot find answers to the questions, it means the visual outputs are not suitable to answer that type of questions and it's not your problem.

The whole experiment involves two stages. The first one is for the test users to individually analyze the visual outputs by looking for information from them to answer specific questions. In the second stage, participants attend a group (so called focus group) to discuss several issues that appeared in the first step. The first stage may take around one and a half hours; the second step will take one hour or so.

It should be specified that video recording will be used in the first step, in order to analyze the details later. The results of the visual analysis will not disclose the names of the test participants.
The test and the focus group will take place between Jan 27 and Jan 30. If you don't have any hesitation, and are willing to join the experiment, please let me know.
Thank you very much.

Kind regards
Qian

APPENDIX 3

Practice questions

Instruction: please use Y/N to indicate whether you can find the answer to the question or not. If it is Y, please write down the answer.

Domain questions	**Year based SOM**		**Station based SOM**			**Co-clustering**
	Cluster map	**Trend plot**	**Cluster map**	**Geographic map**	**Trend plot**	**Geographic map**
1. Which year is the hottest year between 2000-2005 in the whole of the Netherlands?						
2. What is the coldest place in the Southeast of the Netherlands?						

APPENDIX 4

Likert scale Questionnaire

Please give your opinion for the following questions:

1. Do you like the cluster map of year-based and station-based SOM? (It means if it's pleasant to see)

1	2	3	4	5

Dislike Moderate Like

2. Do you like the Trend plot of year-based and station-based SOM?

1	2	3	4	5

Dislike Moderate Like

3. Do you like the Geographical map of station-based SOM and co-clustering?

1	2	3	4	5

Dislike Moderate Like

4. Do you think the Cluster map is of year-based and station-based SOM designed properly?

1	2	3	4	5

No Moderate Yes

5. Do you think the Trend plot of year-based and station-based SOM is designed properly?

1	2	3	4	5

No Moderate Yes

6. Do you think the Geographical map of station-based SOM and co-clustering are designed properly?

1	2	3	4	5

No Moderate Yes

7. Do you think the Cluster map of year-based and station-based SOM is easy to understand? (ease of use)

1	2	3	4	5

No Moderate Yes

8. Do you think the Trend plot of year-based and station-based SOM is easy to understand?

1	2	3	4	5

No Moderate Yes

9. Do you think the Geographical map of station-based SOM and co-clustering is easy to understand?

1	2	3	4	5

No Moderate Yes

APPENDIX 5

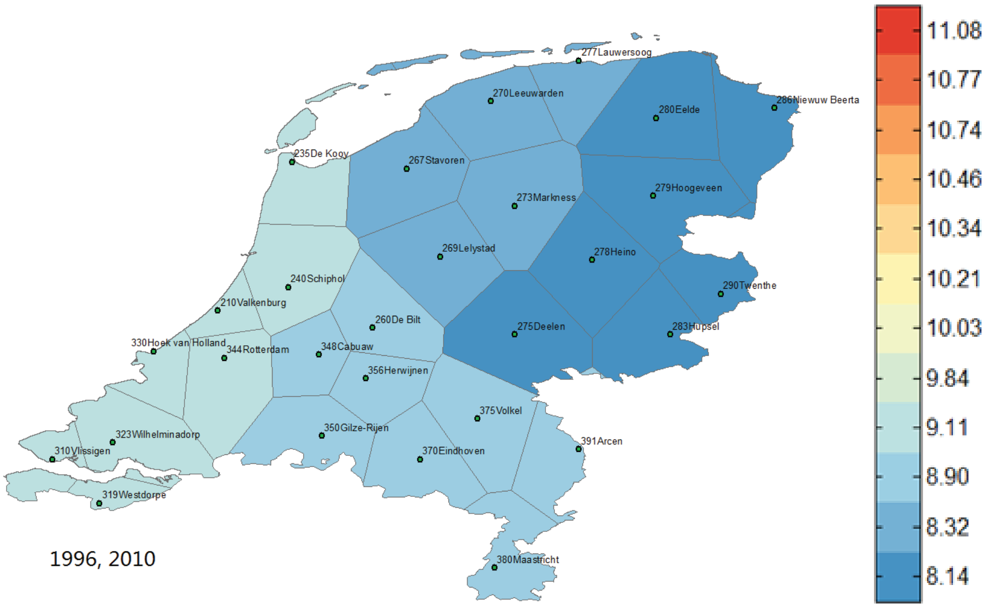
Experiment phase-1 preparation checklist

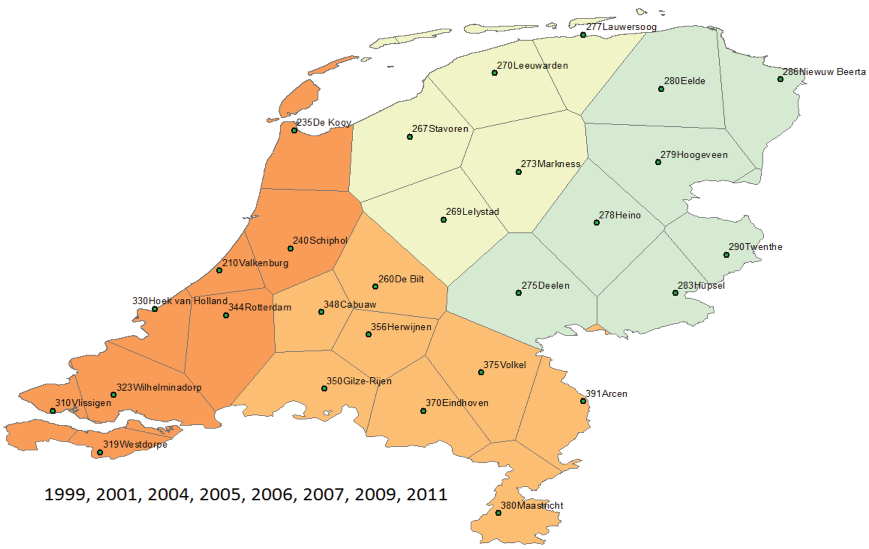
Experiment phase-1 preparation checklist	
Test participants' time schedule	
A laptop	
Video recorder, timer	
Pen, pencil, eraser	
Refreshments	
Paper for taking notes	
Speech to test users	
A participants' information collection form	
Introduction of the clustering theory and the first part of experiment	
A form for users to fill in the answers to the domain questions based on the visual outputs	
Hard copy of clustering visual outputs	
Time cost table	
Likert scale questionnaire	

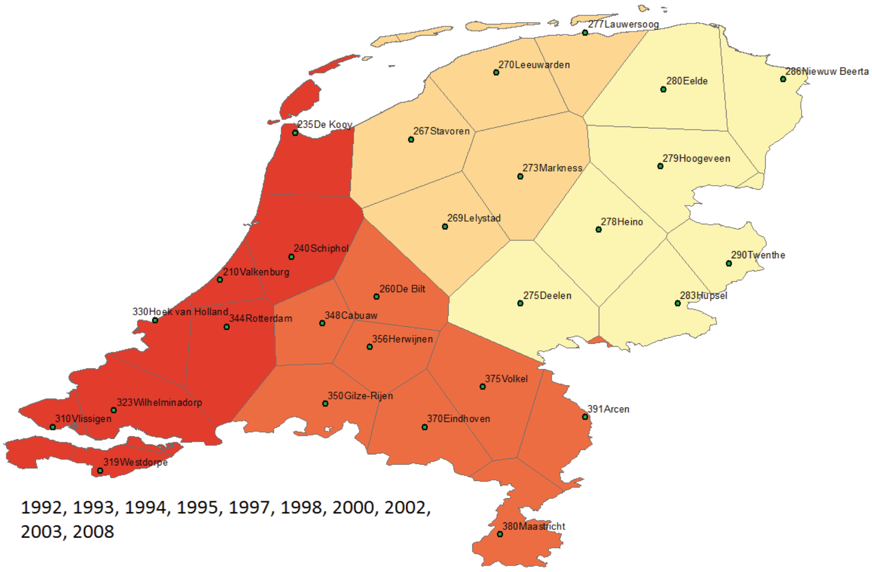
APPENDIX 6-1

Co-clustering geographical map

Illustration: The three graphics together represent the similarity of years and stations simultaneously. each map represents a cluster of years. Inside each year-cluster, the different colors show the different station-clusters. The colors with numbers on the color scale indicates the average annual temperature.



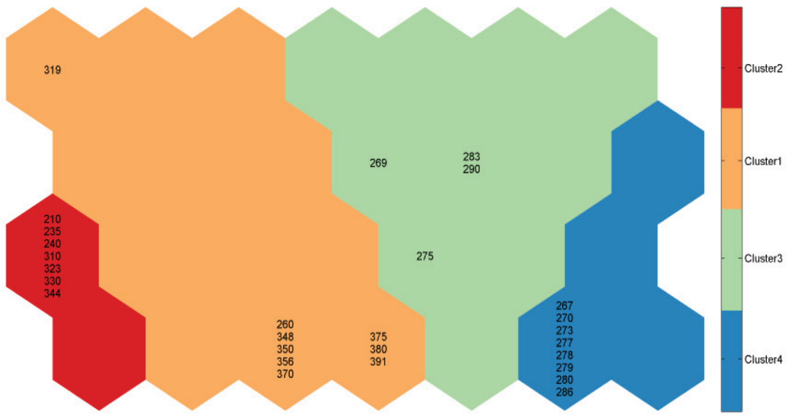



277Lauwersoog
270Leeuwarden
280Eelde
286Niewuw Beerta
235De Kooy
267Stavoren
273Markness
279Hoogeveen
269Lelystad
278Heino
240Schiphol
290Twenthe
210Valkenburg
260De Bilt
275Deelen
283Hupsel
330Hoek van Holland
344Rotterdam
348Cabuaw
356Herwijnen
375Volkel
323Wilhelminadorp
350Gilze-Rijen
391Arcen
310Vlissigen
370Eindhoven
319Westdorpe
380Maastricht
1992, 1993, 1994, 1995, 1997, 1998, 2000, 2002,
2003, 2008

APPENDIX 6-2

Station-based SOM

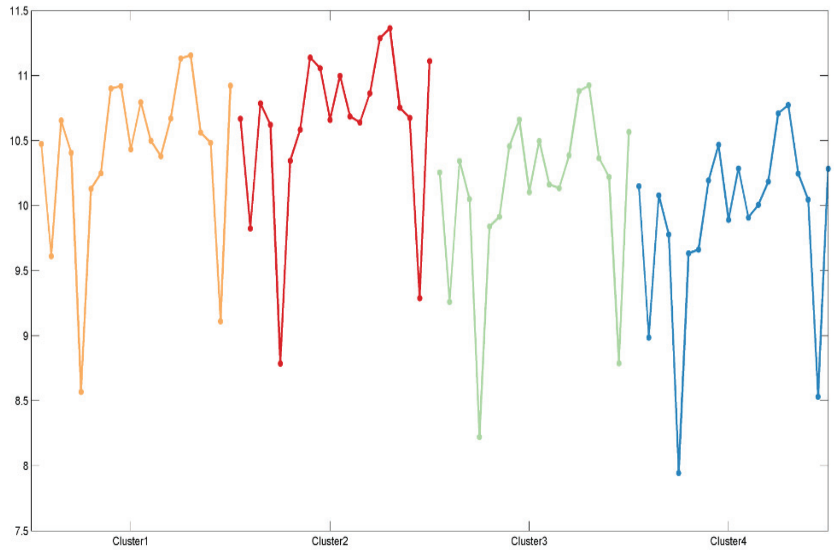
On this page, the four colors show four different station-clusters. The color scale from the top to the bottom indicates the attribute value from high to low.



Graphic 1 Station-based SOM cluster map
This is based on the similarity of 20-year daily temperature data of all the stations.

Graphic 2 Station-based SOM geographic map

The Graphic 2 displays the information of cluster map on a geographic map.



Graphic 3 Station-based SOM trend plot

This trend plot is base on the average annual values of temperature.
The numbers on the Y-axis of the diagram are the average annual values of the temperature.
The four lines, each with 20 points on it, are the 20 years (1992-2011, from the left to the right) of the four station-clusters.
The graphic 3 can be used together with the graphic 1 and 2, as a combination, to answer questions. (Not vice versa)

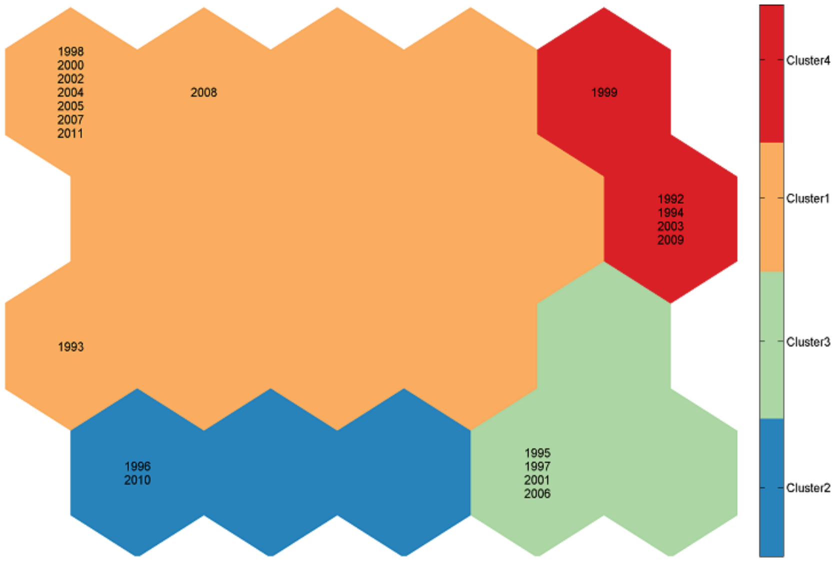
APPENDIX 6-3

Year-based SOM

Illustration

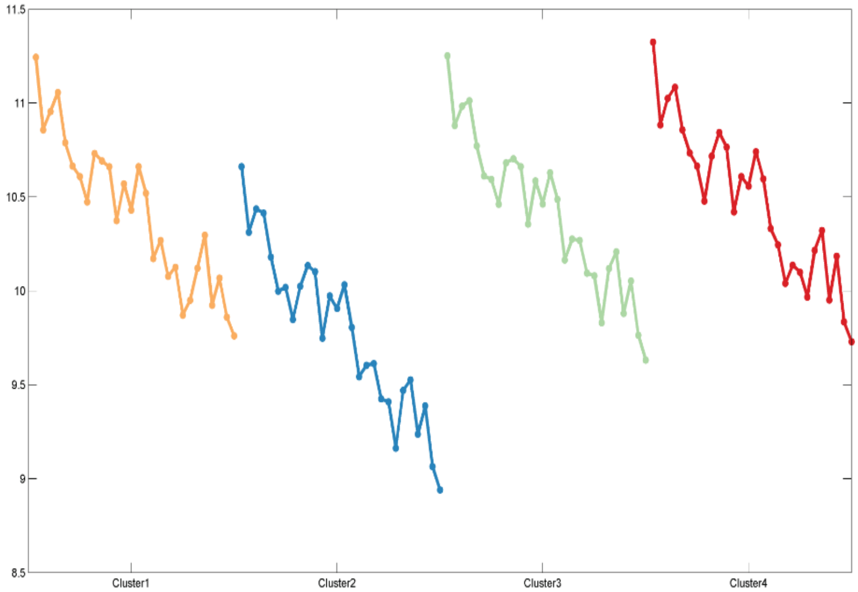
On this page, the four colors show four different year-clusters.

The color scale from the top to the bottom indicates the attribute value from high to low.



Graphic 1 Year-based SOM cluster map

This graphic is based on the similarity of daily temperature data of all the years.



Graphic 2 Year-based SOM trend plot

This trend plot is base on the average annual values of temperature.

The numbers on the Y-axis of the diagram are the average annual values of the temperature.
The four lines, each with 28 points on it, are the 28 stations in the four year-clusters. The 28 points on each line from the left to the right are the 28 meteorology stations from the southwest to the northeast of the Netherlands, station IDs:(310, 319, 323, 330, 344, 210, 350, 348, 240, 380, 370, 356, 260, 375, 391, 235, 275, 269, 267, 273, 278, 270, 283, 277, 279, 290, 280, 286). The graphic 2 can be used together with the graphic 1, as a combination, to answer questions. (Not vice versa)

APPENDIX 7

Speech to test users

Before starting the test

Hi and welcome to join this test, I'm very happy to see you here.

Please take a seat.

First, I will collect your basic information. Please fill in this participants information form (Appendix 5, on my laptop). Then, please read this introduction about the test. (Appendix 6)

Do you have any questions? (I will answer questions, if there are.)

Now, you have five minutes to get familiar with the graphics.

After this, try to finish the two warm-up questions.

Do you have any problem with this?

Shall we start our test?

In the test

Now I will start the video recording. Please inform me when you start each question, and when you finish the test.

Please start answering the questions on the form and thinking aloud in the meanwhile.

Tell me what you are thinking, please keep talking, tell me the things you find confusing, the decisions you are making, etc. (If the test user need to be prompt).

Thanks a lot for your cooperation.

After the test

Please give me your opinion of the questions on this questionnaire.

Around two days away, we will have group meeting for discussing some experience in the test, I will email you the exact time.

Thank you very much for your time.

APPENDIX 8-1

Test Participants' Information

Test user	**Gender**	**Education background (M.Sc / Phd, department)**	**Work experience (what kind of job?)**	**Do you have experience in clustering? (what kind of?)**
P1	Male	M.Sc Geoinformatics B.Sc Geoinformatics	2 years lecturer in university(GIS and Remote sensing)	No
P2	Male	M.Sc. GISc for Urban Planning PGD Land Policy	5 years research in spatial analysis and statistics,	limited - spatial clustering (local Moran's I), k-means
P3	Male	M.Sc. GISc for Natural resources management	9 years in Geoinformmation science	No
P4	Male	M.Sc Geoinformatics, M. Sc HHRR management, B.Sc Software engineering	8 years lecturer in computer sciences, and Spatial Databases	No
P5	female	PhD, ITC Earth Systems' Analysis	Environmental scientist, Assistant researcher	Limited - spatial clustering
P6	male	Cartographic Engineer and PGD in Geoinformatics	Software development in the Brazilian Army	No
P7	female	M.Sc. Natural Hazards and Disaster Riks management	Oil exploration and geodetic surveying	No
P8	male	BSc. and M.Sc. In Water Resources and environmental management	5 year in university teaching (Water related)	No
P9	male	PHD in Big sensor data for smart cities. MSc. In Geoinformatics.	2.5 years lecturer. GIS technician. Agricultural engineer and teacher.	Limited - spatial clustering (local Moran's I) and k-means algorithms

APPENDIX 8-2

Test Participants' Information

Test user	Have you ever analyzed clustering visual outputs like the ones I have just showed you?	Do you have experience with analyzing spatio-temporal datasets?" "If so, what kind of experience?
P1	No	current M.Sc research involves spatiotemporal data extraction and analysis
P2	The geo- map	Spatial-statistical models (with variation mostly on space rather than time)
P3	No	M.Sc research working with spatiotemporal data extraction and analysis
P4	No	In MSc research use time series data about land usage change and land ownerships
P5	No	Professional experience with environmental data. MSc and PhD-related experience with remote sensing data
P6	No	No
P7	No	In MSc reseach working with time series made of thermal infrared datasets
P8	No	In MSc reseach working with time series landsat images
P9	No	Visualization and representation of uncertainty in moving object datasets.

APPENDIX 9

Introduction of clustering technique and the experiment

Introduction of clustering:
The clustering technique is an exploratory data analysis tool that can be used to discover structure and regularity in big dataset. After clustering, the objects with similar attributes are grouped together in the same cluster.

Introduction of the task:
The case study dataset: the Netherlands 20 years daily temperature data, from 28 meteorological stations.
In this task, the participant will try to find answers to the domain questions in the visual outputs provided. The domain questions are about the case study dataset, the visual outputs are produced by the two clustering techniques based on the same case study dataset.

In front of you, there is a laptop for exhibiting the form of the domain questions. On the table, there are three hard copies of the graphics that are going to be tested. They are visual outputs of the three ways of clustering (Year-based SOM, Station-based SOM and co-clustering. Year-based SOM and Station-based SOM mention the dataset is grouped from the time aspect and location aspect respectively.) You need to try to find answers to the domain questions in the provided visual outputs and fill in the form.
You are kindly required to go through all the graphics on three hard copies for each question, which means fill in the blanks by row.

When performing the task, you are requested to think aloud, which means speak the thoughts in your mind. The thoughts are only about the test, not about the secret thinking.
In the process of the task execution, you are encouraged to propose any questions you have about this task. But as an observer, I will only record the questions and will not respond.

Whenever you feel tired, we will take a break; whenever you want to quit the test, we will stop. Each test participant will be given five minutes to get familiar with the visual outputs before executing the task and two test questions as a practice.

APPENDIX 10

Test scenario

Before starting the test:
Arrive at the laboratory half an hour earlier than the test users. Set out the refreshments.
Start my laptop, open:
Test participants' information form with another row hided) & time cost table
Put the hard copies of graphics on the table.
Start the user's laptop:
Open the PPT.
Open the DQ document, the warm-up question, and then minimize it.
When the test user come in:
Welcome to join my pilot study.
First, I will collect your basic information. Please fill in this participants information form (on my laptop).
Then switch to the time cost table .
Thanks. Now, we will start with a PPT. It's about the basic theory of this test and task introduction.
Do you have any questions? (I will answer questions, if there are.)
Then, I'd like to show you a short demo about thinking aloud.
Now, you have five minutes to get familiar with the graphics.
After this, try to finish the three warm-up questions.
Do you have any problem with this?
Shall we start our test?

In the test
Now I will start the video recording. Please inform me when you start each question, and when you finish the test.
Please start answering the questions on the form and don't forget thinking aloud in the meanwhile.
If the test user needs to be prompt, user sentence like, tell me what you are thinking, please keep talking, tell me the things you find confusing, the decisions you are making, etc.
Thanks a lot for your cooperation.
After the test
Please give me your opinion of the questions on this questionnaire.
Around two days away, we will have a group meeting for discussing some experience in the test, I will email you the exact time.
Thank you very much for your time.

APPENDIX 11

Introduction and illustration of the test- PowerPoint

Slide 1
Introduction 1 Basic Theory
Key words
Big spatiotemporal datasets
Clustering
Visual outputs of clustering
The clustering technique is an exploratory data analysis tool that can be used to discover structure and regularity in big datasets.
After clustering, the objects with similar attributes/behavior are grouped together in the same cluster.
Visualization of the clustering results may help to interpret the data.

Slide 2
Introduction 2 Test
In this research study, we have a big spatiotemporal dataset about temperature in the Netherlands from 28 meteorological stations (daily temperatures for a period of 20 years (1992-2011)).
The clustering is based on the similarity of original daily temperature data.
Two clustering methods (SOM & co-clustering) were applied to the original data. the visual outputs that resulted from these clustering are grouped into three categories: Year-based SOM, Station-based SOM and co-clustering.

Slide 3
Introduction 3 Examples of visual outputs- station-based SOM (Exactly the same as the appendix 6-2)

Slide 4
Introduction 4 Think aloud method
In this test, you are asked to execute a number of tasks with visual outputs like the ones you have just seen.
We ask you to THINK ALOUD while doing so.
Thinking aloud means to speak out all the thoughts about the test that come up in your mind, when you are performing the test.
You are also encouraged to speak out any questions you may have about this task. However, as an observer, I will only record the questions and will not respond.
The questions can be like:
I'm wondering, I'm confused about, this remind me, I'm not sure, what/why/where is, etc.

Slide 5

Introduction 5 Instructions
Your task is to try to find out which visual outputs can offer what kind of information, by answering pre-defined questions (showing on the screen) about this dataset. The answers are suggested to write in a simple way. For example, the answer is the content of the blue cluster.
You need to use the provided visual outputs to answer questions on the form, and fill in the answers in the form on the computer. First answer all the questions for one group of visual outputs, then for the next group and then for the third group.
You need to start each question by saying, now I'm starting question 1, 2, …
Whenever you feel tired, we will take a break; whenever you want to quit the test, we will stop.

APPENDIX 12-1

Time cost table-Year based SOM

Graphic	Year based SOM																	
	Cluster map									Trend plot								
Domain question	p1	p2	p3	p4	p5	p6	p7	p8	p9	p1	p2	p3	p4	p5	p6	p7	p8	p9
Q1	27	17	7	7	5	5	11	10	9	18	17	51	14	38	45	21	51	32
Q2	18	22	12	6	4	12	5	4	8	10	31	10	40	14	7	18	31	19
Q3	10	12	15	8	4	11	6	5	6	9	29	12	151	23	12	9	35	58
Q4	12	11	11	4	4	4	5	5	8	9	9	9	160	34	18	6	12	12
Q5	13	9	8	6	4	6	7	15	6	120	10	9	24	5	7	14	32	21
Q6	9	6	12	5	5	4	5	7	4	13	11	15	32	7	4	43	9	42
Q7	10	16	9	8	3	5	5	5	9	90	27	13	21	6	33	14	31	32
Q8	14	10	8	14	7	18	12	5	17	67	12	11	22	18	6	12	14	11
Q9	8	40	7	6	6	7	11	7	18	8	15	9	28	16	40	53	24	97
Q10	7	15	11	8	4	3	11	11	8	12	19	15	75	6	6	7	7	86
Q11	16	30	8	15	8	4	8	10	13	7	12	5	17	158	8	9	39	115
Q12	6	12	9	5	7	3	11	5	14	23	17	11	61	36	5	93	96	66
Q13	14	7	31	7	5	3	6	13	12	8	20	12	38	21	9	5	16	32
Q14	9	8	6	7	5	4	45	41	16	7	41	11	16	42	11	16	19	17

Time cost table-Station based SOM

Graphic Domain Question	Station based SOM																	
	Cluster map									Geographic map								
	p1	p2	p3	p4	p5	p6	p7	p8	p9	p1	p2	p3	p4	p5	p6	p7	p8	p9
Q1	9	9	11	6	32	5	11	16	16	61	10	9	7	2	4	8	14	18
Q2	12	12	8	9	5	4	9	22	18	15	11	4	4	7	3	6	27	13
Q3	11	22	12	48	6	10	11	10	22	14	16	7	41	8	4	17	8	20
Q4	33	9	5	9	5	8	21	11	8	33	8	3	13	4	2	12	54	5
Q5	15	10	7	14	3	7	6	6	8	11	6	4	12	2	6	6	6	6
Q6	20	8	9	4	3	5	5	2	9	10	4	3	5	2	4	5	4	4
Q7	6	8	7	4	3	4	9	5	14	7	7	3	4	2	5	6	5	6
Q8	11	9	6	11	5	7	9	7	8	13	9	3	14	4	8	15	5	15
Q9	8	11	11	6	7	9	22	28	25	14	8	6	8	4	5	14	13	11
Q10	12	12	15	51	9	9	11	7	11	7	9	12	50	2	6	17	6	18
Q11	21	11	16	9	3	6	13	6	10	9	7	4	8	3	4	8	5	13
Q12	11	24	8	14	2	7	10	5	10	19	6	3	26	3	4	9	5	14
Q13	21	8	7	5	3	4	9	6	20	21	7	9	9	5	8	8	26	14
Q14	8	7	5	8	3	3	13	48	32	23	6	3	17	9	11	36	14	26

Time cost table-Station based SOM trend plot

Graphic Domain Question	Station based SOM								
	Trend plot								
	p1	p2	p3	p4	p5	p6	p7	p8	p9
Q1	34	12	18	31	42	21	15	71	42
Q2	41	26	5	34	4	5	6	68	32
Q3	38	18	31	59	19	21	11	12	18
Q4	24	7	6	16	4	12	8	8	4
Q5	190	14	12	11	56	32	32	42	120
Q6	22	12	5	32	4	9	11	21	41
Q7	33	11	4	18	6	31	4	22	61
Q8	58	12	8	32	31	18	14	12	42
Q9	96	29	21	49	7	8	12	49	44
Q10	29	21	9	52	52	16	11	32	23
Q11	61	51	13	42	41	28	66	53	51
Q12	33	14	9	12	18	15	13	20	28
Q13	21	8	5	41	18	16	6	18	27
Q14	24	16	7	18	26	9	16	10	22

Time cost table- Co-clustering

Graphic Domain Question	Co-clustering								
	Geographic map								
	p1	p2	p3	p4	p5	p6	p7	p8	p9
Q1	75	15	51	56	44	11	20	63	48
Q2	40	20	22	18	16	12	15	32	41
Q3	31	23	47	16	21	21	10	11	13
Q4	41	38	42	23	14	18	11	6	26
Q5	20	24	41	11	18	12	28	28	22
Q6	18	10	16	17	8	10	22	22	14
Q7	15	22	19	32	9	10	50	33	15
Q8	13	10	17	7	7	9	12	9	9
Q9	44	56	19	8	59	5	22	36	41
Q10	14	22	21	11	53	31	48	8	47
Q11	12	32	33	23	36	22	19	32	52
Q12	10	25	30	61	20	14	23	34	51
Q13	23	16	32	11	11	14	26	23	14
Q14	11	34	17	26	38	15	42	20	28