Aalto University
School of Engineering

Exploring the context of maritime search and rescue missions using visual data mining techniques

Master's Thesis, Department of Surveying and Planning, School of Engineering, Aalto University

Espoo, 30. April 2014

Bachelor of Science in Technology
Maikki Sonninen

Supervisor: Professor Kirsi Virrantaus
Instructors: Dr. Paula Ahonen-Rainio,
M.Sc. Floris Goerlandt

Aalto University
School of Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Maikki Katrina Sonninen

Title of thesis Exploring the context of maritime search and rescue missions using visual data mining techniques

Degree programme Degree Programme in Geomatics

Major/minor Geoinformatics **Code of professorship** Maa-123

Thesis supervisor Professor Kirsi Virrantaus

Thesis advisor(s) Dr. Paula Ahonen-Rainio & M.Sc. Floris Goerlandt

Date 30.04.2014 **Number of pages** 126 **Language** English

Abstract

Recreational boating in Finland is a popular activity. The environment of the Gulf of Finland attracts boaters and other tourists to spend time at sea. The increase of boating activity raises the potential number of boating accidents which may be caused by several factors. The Finnish Border Guard is the leading authority of Maritime SAR operations. Other authorities, mainly the Finnish Rescue Services, may attend maritime SAR operations if necessary. Volunteer organizations, such as the Finnish Lifeboat Institution, may also participate in maritime rescue activity.

The goal of this master's thesis was to apply visual data mining techniques and compare their suitability in finding relationships between the SAR incidents and weather parameters. The comparison of the activity of SAR organizations are also included as part of the analysis. The study aims to define under what kinds of weather conditions certain types of incidents tend to happen and which SAR organization is taking action most often during challenging conditions. The study is focused on the Gulf of Finland.

The study starts with a literature review of Maritime Search and Rescue system and boating safety in Finland in general. The literature review continues with a data mining topic focusing on the context of multidimensional data and visualization techniques applied in this study. Description of the data and data processing is included thereafter.

In the research, it is found that most incidents assigned with challenging weather conditions were located in the western Gulf of Finland. Some incidents were noticed in the East as well. The metropolitan area was assigned to have the most challenging wave heights at one buoy in the Gulf of Finland. The most interesting weather parameters were wind speed and wave height.

It was also noticed that the occurrence of studied incident types are random and every SAR organization operates missions despite the weather or time of the day. Some differences in travel distances of SAR organizations were noticed.

A more detailed analysis was applied to the data of the Finnish Lifeboat Institution. In this part, information on accident factors, boat types and sizes was included. Also an example of comparing SAR vessels was made.

Keywords Visual data mining, multidimensional data, maritime search and rescue, recreational boating, boating safety

Aalto-yliopisto
Insinööritieteiden korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Diplomityön tiivistelmä

Tekijä Maikki Katrina Sonninen	
Työn nimi Meripelastustehtävien tutkiminen käyttäen visuaalisia tiedonlouhintamenetelmiä	
Koulutusohjelma Geomatiikka	
Pää-/sivuaine Geoinformatiikka	**Professuurikoodi** Maa-123
Työn valvoja Professori Kirsi Virrantaus	
Työn ohjaaja(t) TkT Paula Ahonen-Rainio & Dipl.Ins. Floris Goerlandt	
Päivämäärä 30.04.2014 **Sivumäärä** 126	**Kieli** englanti

Tiivistelmä

Suomessa huviveneilyn suosio on suuri. Suomenlahden ympäristö houkuttaa veneilijöitä ja muita matkailijoita viettämään yhä enemmän aikaa vesillä. Veneilyharrastuksen kasvu samalla nostaa veneilyonnettomuuksien lukumäärää, mikä voi olla usean tekijän summa. Rajavartiolaitos on Suomen johtava meripelastustoimen viranomainen. Jos katsotaan tarpeelliseksi, muut viranomaiset, pääsääntöisesti Pelastuslaitos, voivat osallistua meripelastustoimintaan. Vapaaehtoisjärjestöt, kuten Suomen Meripelastusseura, voivat myös suorittaa meripelastustehtäviä.

Tämän diplomityön tavoite oli käyttää visuaalisen tiedonlouhinnan menetelmiä ja verrata niiden sopivuutta meripelaustapahtumien ja säämuuttujien välisten suhteiden tutkimiseen. Eri meripelastusjärjestöjen toiminnan vertailu on myös osa analyysia. Tutkimuksen tavoitteena on määrittää, minkälaisissa olosuhteissa tietynlaiset meripelastustehtävät yleensä ilmenevät ja mikä meripelastusjärjestö on aktiivisin vaikeissa olosuhteissa. Tutkimus on kohdistunut Suomenlahteen.

Työ alkaa kirjallisuuskatsauksella, joka käsittelee meripelastusjärjestelmää ja veneilyturvallisuutta Suomessa. Kirjallisuuskatsaus jatkuu visuaalisen tiedonlouhinnan aiheella, jossa tarkastellaan moniulotteista dataa ja tutkimuksessa käytettyjä visualisointitekniikoita. Sen jälkeen kuvaillaan tutkimuksessa käytettyä dataa ja sen käsittelyä.

Tutkimuksessa havaittiin, että useimmat haastavissa sääolosuhteissa tapahtuneet veneonnettomuustyypit sijoittuivat läntiselle Suomenlahdelle, mutta myös jonkin verran itään. Kovimmat tuulen nopeudet havaittiin myös näillä alueilla. Pääkaupunkiseudun alueella oli tunnistettavissa haastavimmat aallonkorkeudet.

Tutkimuksessa huomattiin myös, että tutkitut meripelaustehtävätyypit ilmenevät satunnaisesti ja että meripelastusjärjestöt pysyvät aktiivisina säästä ja ajankohdasta huolimatta. Joitakin eroavaisuuksia oli huomattavissa eri meripelastusjärjestöjen matkojen etäisyyksissä.

Suomen Meripelastusseuran tuottamalle datalle suoritettiin tarkempi analyysi, jossa tarkasteltiin onnettomuustekijöitä, venetyyppejä ja venekokoja. Työssä tehtiin esimerkkinä vertailu pelastusaluksista.

Avainsanat Visuaalinen tiedonlouhinta, moniulotteinen data, meripelastus, huviveneily, veneilyturvallisuus

Preface

Before I started my freshman year in Helsinki University of Technology, I did not expect to end up studying Geoinformatics and Cartography as my major. The field of study sounded new and peculiar that I did not know much about. The student life has been quite a journey and now it is time for new adventures after seven years of studying. I am looking forward to seeing where I am going to end up next and what opportunities the future will provide me.

I would like to thank my instructor, Paula Ahonen-Rainio, for instructing me and for giving constructive advice and feedback. Thanks also belong to my other instructor, Floris Goerlandt, for his feedback, advice and for letting me be part of RescOp project. I would like to thank Kirsi Virrantaus for supervising me. Finally, my thanks belong to Mom and Dad for giving me their support and cheering me up all this time.

Whilst doing and writing this Master’s Thesis, I’ve been having quite a journey already. I hope all this, what I’ve exceeded and experienced so far, will assist me make decisions among all the opportunities the world provides me. I believe every single day will bring me a step closer to my dreams.

Table of Contents

Figures

Tables

Abbreviations and terms

Bivariate map	Cartographic technique of mapping two attributes in one map.
FBG	*Finnish Border Guard*. The national security agency of Finland responsible for security of borders of Finland at land and sea (Rajavartiolaitos 2014).
FLI	*Finnish Lifaboat Iinstitution*. The umbrella organization for voluntary maritime rescue associations in Finland (Suomen Meripelastusseura 2014).
FMI	*Finnish Meteorological Institute*. A research and service agency under the Ministry of Transport and Communications responsible for compiling and providing atmospheric data and forecasts in Finland.
FRS	*Finnish Rescue Services*. Rescue services in Finland that are led by the ministry of Interior's Rescue Department. (City of Helsinki 2013.)
GeoMap	A classified bivariate choropleth map. A component of GeoViz Toolkit.
GeoViz Toolkit	An application version of GeoVista Studio. An open-source, Internet-delivered program for geographic visualization and analysis (Hardisty & Robinson 2009, p. 191; available: http://www.geovista.psu.edu/geoviztoolkit/).
KKJ	*Kartastokoordinaattijärjestelmä*. A National Grid Coordinate System of Finland (Finnish Geodetic Institute 2012).
Maritime SAR	*Maritime Search and Rescue.* Service to assist people in distress or danger at sea (Rajavartiolaitos 2014).
MESSI	The statistics of rescue resources and accidents compiled and maintained by the Finnish Lifeboat Institution.
Multidimensional data	Data that consist of several attributes.

Multiform Bivariate Matrix	A generalization of a scatterplot matrix consisting of univariate visualizations.
PCP	*Parallel Coordinate Plot*. A data visualization technique used for plotting multivariate data. The data rows of a data table are represented as polylines. Each attribute of a row is represented by a point on the line.
PELA	*Pelastuslaitos*. Finnish term of the Finnish Rescue Services.
PRONTO	*Pelastustoimen resurssi- ja onnettomuustilasto*. The statistics of rescue resources and accidents compiled and maintained by the Emergency Services College.
RescOp	A 3-year Finnish-Russian project focusing on development of rescue operations in the Gulf of Finland (KMRC 2014).
RVL	*Rajavartiolaitos.* Finnish term of the Finnish Border Guard.
RVT	*Rajavartiotoiminnan tietojärjestelmä*. The statistics of rescue resources and accidents compiled and maintained by the Finnish Border Guard.
SAR	*Search and Rescue*. An emergency service involving the detection and rescue of people in accident or in danger.
Scatterplot	A graph of plotted points representing the correlation between two attributes.
Scatterplot matrix	A matrix consisting of grid cells of scatterplots.
SMPS	*Suomen Meripelastusseura*. Finnish term of the Finnish Lifeboat Institution.
SRU	*Search and Rescue Unit*. A vessel or aircraft with a qualified and equipped crew to perform SAR operations.
WGS84	*The World Geodetic System 1984*. A reference frame for the earth for use in geodesy and navigation (NGA 2013).

1 Introduction

1.1 Background

The environment of the Gulf of Finland is an exceptional area where hard winter conditions, heavy maritime traffic and international affairs meet. This increases the vulnerability of the region to accidents at sea. Recreational boating is a popular activity in Finland during the summer months, and the popularity and services of boating are constantly increasing. Improving the maritime safety requires preparations for environmental risks and other threats that the increasing traffic flow may cause. The increase of recreational boating and traffic increases the need of maritime search and rescue missions now and in the future.

Maritime search and rescue means rescuing and safeguarding human lives in occasions of emergency and danger situations, but also rescuing of material, such as vessels and cargo. The Maritime Search and Rescue in Finland is led and controlled by the Ministry of the Interior. The Finnish Border Guard (FBG) is the leading authority of the Maritime Search and Rescue prescribed. Other authorities, mainly the Finnish Rescue Services (FRS) may participate in the activity if it is necessary. Voluntary rescue organizations, such as the Finnish Lifeboat Institution (FLI) can perform rescue missions that are approved by an official rescue authority. The maritime safety in the Gulf of Finland is aimed to be improved through cooperation between these SAR organizations. Although the maritime safety has been enhanced, there still is need for gaining more knowledge of the risk factors and the environment. (Räsänen et al. 2005, p. 48.)

This study was initiated by RescOp project, a 3-year Finnish-Russian project on development of rescue operations in the Gulf of Finland funded by the European Union, Russian Federation and the Finnish State (ENPI CBC). The main focus of the project is on the improvement of maritime safety and reducing environmental risks through cross-border cooperation, research, development and training. The goal of the project is to develop tools for risk management for the main environmental challenges.

An atlas about the SAR incidents in Finland (Venäläinen & Sonninen 2013) analyses the Finnish maritime and lacustrine SAR missions and responses during the years 2007–2012. The focus was on the missions performed by the FLI but also performances of the FBG and the FRS were included. The atlas provides a visual overview of Finnish water traffic accidents in the sea and lacustrine areas. The target of this thesis is to bring the analysis forward and deeper by additionally considering additional weather and wave data using visual exploration techniques. The aim of this research is to compare missions and the level of activity of each SAR organization.

1.2 Visual data mining of multidimensional data

Exploring and analyzing data related to space and time is a challenging task, especially when it comes to a large set of attributes. Multidimensional datasets tend to be high in volume, which easily causes tool limitation issues, for example, the tools of today are not advanced enough. Today's technology enables computational systems to store large amounts of high-dimensional data, but can lack in providing solutions to deal with both spatial and temporal dimensions. Human vision and cognition in perceiving trends and patterns from a complex dataset are limited. For this reason, interesting subsets of variables have to be formed in order to discover the relationships behind the given dataset, although the data might already be selected according to a type of characteristics, for example, numeric or non-numeric, spatial or non-spatial, linear or non-linear, etc. (Chen et al. 2008, p. 33; Guo et al. 2005, p. 113).

Thomas and Cook (2006, p. 10) define visual analytics as the science of analytical reasoning supported by interactive visual interfaces. People are using more and more visual tools and techniques to get insights from massive datasets that include dynamic, heterogeneous and conflicting data. This way they can detect the expected and find unexpected patterns. In order to do so, the user needs to get an access to the tools and techniques.

Shneiderman (1996, p. 337) who was presenting guidelines for creating a user interface for an interactive information visualization tool, proposed that an effective tool for visualization should function along the principles of the mantra: overview first, zoom and filter, and detail-on-demand. This is called the Information Seeking Mantra. The mantra describes an approach where the user first gains an overview of the data set. The next step is to search interesting patterns and drill down into detailed information analysis by using tools for zooming and panning, and by using data filtering and extraction for data reduction. For discovering more details of the subsets, the user needs to be capable to get an access to the details by using suitable techniques. The tasks of information visualization have been further described by Siirtola and Räihä (2006, p. 1280) who discuss about parallel coordinate features in their study. The task taxonomy includes:

Overview: Construct an overview of the whole data set.
Zoom: Zoom in on interesting patterns.
Filter: Filter out irrelevant items.
Details-on-demand: Select an item or a group of items for detecting details.
Relate: Study relationships between the items.
History: Keep a history on actions to support undo, replay, and refinement.
Extract: Allow extraction of sub-collection and of the parameters used in the query.

1.3 Objectives and research questions

The target of this study is to generate understanding and knowledge for the need of SAR operations through visual analysis of existing data. The approach in this study is a data mining method using typical visual analysis techniques. The interest of the study is in the outliers and possible dependencies between the variables. The applied techniques will be compared and evaluated according to their suitability of the exploration.

1.3.1 Research questions

The goal of the research is to find answers to the following research questions:

- What are the general conditions when a certain type of an incident tends to happen?
- Under what conditions do the SAR organizations perform their missions? Are there any difference?
- Which visual analysis technique is suitable in discovering the patterns? Can we find any dependencies or outliers?

In the first question, the focus is on the perspective of the incident where both the temporal and spatial occurrence is relevant. The second question compares the activity of the SAR stations within the time period 2007–2012. The general-level analysis concerns the data of all SAR organizations together. The more detailed analysis focuses on the data of the FLI.

1.3.2 Data

The same SAR data that was used in the Atlas (Venäläinen & Sonninen 2013) is recorded from missions performed during the years 2007–2012. Now the data includes additional weather observation data and wave data that are linked to the incidents. In addition, observations of lightning strikes are also included to the data set.

Altogether, the data in this research consists of 5 411 cases. The major targets of the missions are recreational vessels. Some vessels are assigned as fishing vessels (12) and vessels that are not clearly defined as recreational or other (467). The incident data derive from three different SAR organizations, the Finnish Border Guard (FBG), the Finnish Lifeboat Institution (FLI), and the Finnish Rescue Services (FRS). The focus is concentrated on the main boating season which is during the summer months, when each SAR organization is mostly active (Figure 1).

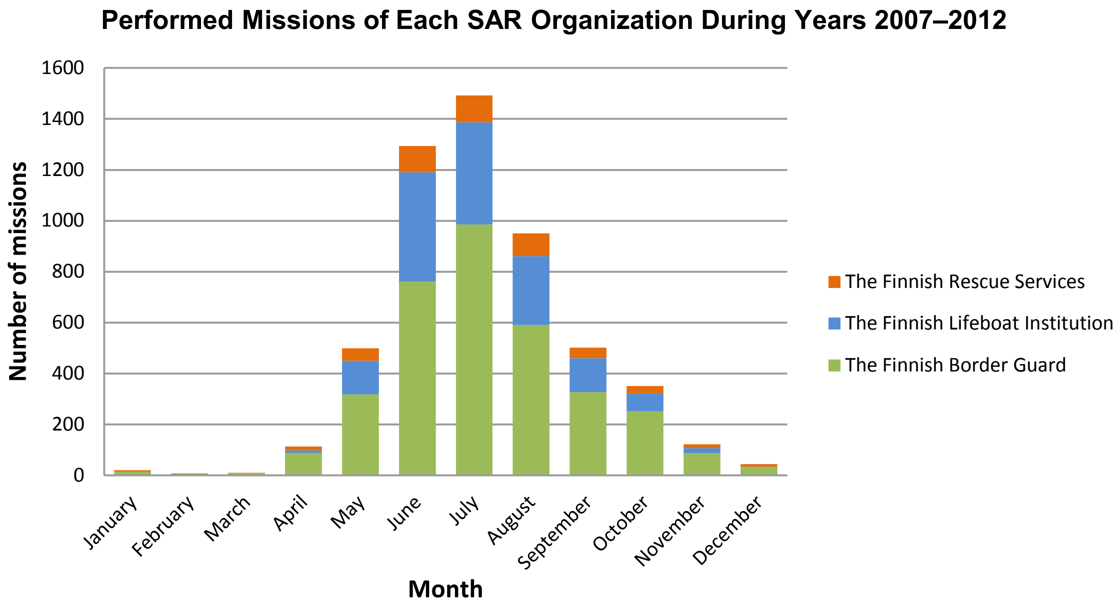


Figure 1 *Diagram showing the distribution of incidents of each SAR organization in the Gulf of Finland.*

Figure 2 shows that the FLI is mainly active during the boating season. The other organizations are active throughout the year. This is taken into account and the task of the analysis is to compare the overall conditions during the time periods when all the organizations are on duty. It is also known that the majority of the missions are performed by the FBG (3 474). Although, the FRS is active throughout the year, the performance of the incident is the minority (467). The main explanation is that the maritime search and rescue is assigned primarily to the FBG and the FLI.

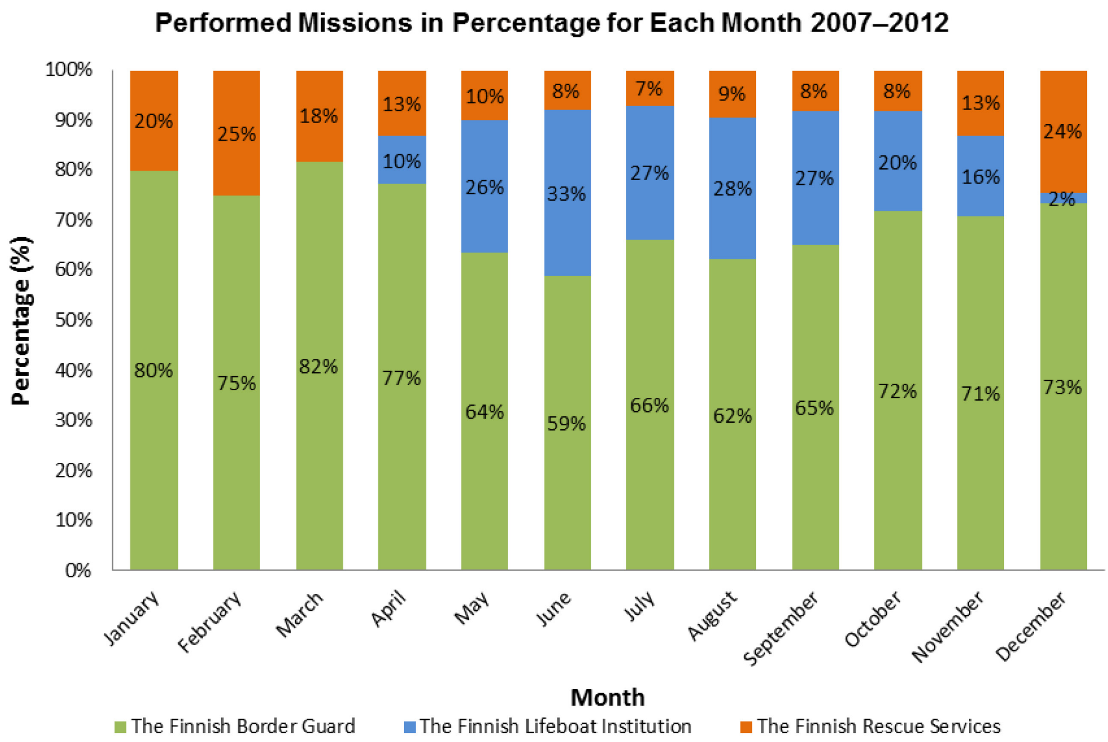


Figure 2 *Missions shared by SAR organizations.*

The more detailed analysis focuses on the data compiled by the FLI, which consists of 1 470 recorded cases in the Gulf of Finland. The reason for this lies in the consistency of the data as well as the Atlas, which is more focused on the incidents performed by the FLI. In the more detailed analysis, the type and size of the vessels as well as the SAR vessels taking action on the missions will be included in the analysis. Also accident qualities and accident factors are explored.

The aim of this study is to investigate the possible impacts of one variable to another also using secondary data. The distances between the incident and the SAR station and the shore have been calculated separately. Travel times, operation times and end times are not provided. The received alarm time is given for every incident and is thus assumed to be the time when the incident happened.

1.3.3 Limitations

RescOp project is focused on the maritime safety in the Gulf of Finland area. Thus the incidents included in this analysis are placed in the Gulf of Finland waters. The target vessels are mostly recreational boats. Other vessel types, such as cargo ships, cruisers and other vessels that are clearly not recreational vessels, are not included in the analysis. The observed area is West-East oriented starting from Hanko (West) to the Russian territorial area (East). This limitation is defined according to the Finland's Search and Rescue Region (SRR).

The SRR is divided into two sub-regions, the West Finland Coast Guard Sub-Region coordinated by the Turku Maritime Rescue Coordination Centre (MRCC Turku) and the Gulf of Finland Coast Guard District coordinated by the Helsinki Maritime Rescue Sub-

Centre (MRSC Helsinki) (Figure 3). This research concentrates on the MRSC Helsinki only.

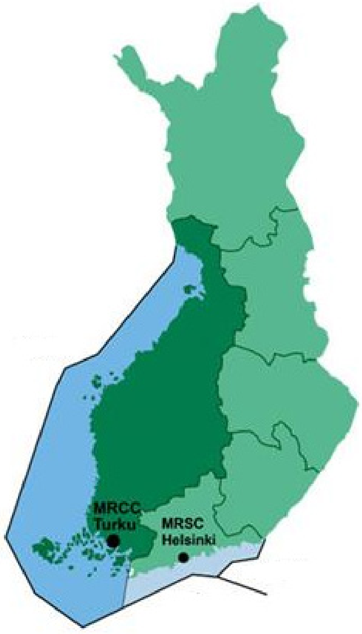


Figure 3 *(edited) Division of the Search and Rescue Region into two sub-regions. (Rajavartiolaitos 2014). Source: http://www.raja.fi/download/00325994dc5646b.jpg. [Referred 13.1.2013].*

This research furthermore only considers recreational vessels, which include normal boats, motorboats, sailboats and other vessels that are used for recreational purposes, such as jet-skies, canoes, kayaks, sail dinghies and boards. The vessel types are classified as sailboat, motorboats and other vessel types. The additional weather and wave data are applied so that it would cover the area as evenly as possible.

1.4 Methods

The analysis was done by applying the techniques that are implemented in GeoViz Toolkit, which is a set of tools developed for the visualizing multidimensional geographic data (Hardisty & Robinson 2009). It has the commonly used components that support analysis of multivariate data in a geographic space. The environment and characteristics of the applied techniques are described in Appendix 6. (GeoViz Toolkit available: http://www.geovista.psu.edu/geoviztoolkit/).

It is obvious that the work layout had to be according what the toolkit provides. The environment and techniques are already implemented in the package. The views are dynamically linked to each other providing mapping and statistical graphing components for univariate and multivariate data. The applied techniques are provided with the Geoviz Toolkit. Literature review of the techniques is provided, which is the basis for the analysis. The process included data modification which is described in its own chapter.

The task is not to study trivial issues or assumptions. The main focus is on outliers, differences or similarities between the stations and incidents, and finding interesting patterns, such as dependencies.

1.5 Structure

The system of maritime search and rescue, boating safety and maritime SAR-related research are introduced in chapter 2. Chapter 3 describes the basic concept of the visual data mining approach and describes the visual techniques applied in this research. The techniques are typical for visual exploration: parallel coordinate plot (PCP), Multiform Bivariate Matrix and bivariate map.

Chapter 4 describes the data, data preprocessing including the classification criteria applied to the data. Chapter 5 presents the results that the techniques provided and compares the SAR organizations. Chapter 6 discusses the final results, and chapter 7 lists the conclusions.

2 Maritime search and rescue in Finland

This chapter introduces the boating safety in Finland and boating-related research on causes of boating accidents that have been studied. The Finnish maritime SAR system is presented. A review of the risk-informed decision making for SAR location planning is introduced at the end of the chapter.

2.1 Boating and safety in Finland

In the broad sense, recreational boating means all types of leisure activity that is involved with a boat, canoe, jet-ski or any kind of water vehicle (Räsänen et al. 2005, p.13). The development and the economic growth of our society have increased the leisure time. Simultaneously, recreational boating and living along waterways have increased significantly. Finnish lakes and sea coasts provide an ideal environment for recreational boating. Finns spend lots of time in summer cottages, which most of them are located either at a lake or on a sea coast. The Finnish waters provide an opportunity for several different water activities such as sailing, motor boating, canoeing, jet-skiing, fishing, row boating, etc. New boating-related trends have become more popular, which has increased the concern on boating safety. (Luntiala et al. 2004, p. 1; Mäkilä 2009, p. 8.)

According to a study (Luntiala et al. 2004, pp. 8 & 12), boating accidents in Finland occurred more frequently on lakes (74 %) than in the seas (26%) during the years 2000−2002. In the lake areas, the most accidents happened in small waters rather than in bigger lakes. In the study, the boat types were categorized into sailboats, motorboats and rowboats. The study showed that the number of incidents has decreased, but the major group of accident was still open rowboats equipped with or without a small engine. Accidents with jet-skies as a new trend were very rare in the statistics, but they were claimed to be both severe and destructive due to their high speed.

Registration of motored recreational boats is mandatory in Finland. Almost all sailing yachts are included in the register of the Finnish Yachting Association. Smaller boats, dinghies, canoes, kayaks, rowboats and jet-skies are not in any register. Around 20 % of the total number of motorboats in Finland is located on the eastern Gulf of Finland. The highest number is in the regions of Helsinki, Porvoo and Kotka. (Deltamarin 2006, p. 5.)

The trend of high power motor vessels and inexperienced boating skills bring new challenges. Collisions and groundings with boats traveling with high speeds can be like car accidents. Grounding is the most common damage type in the Finnish seas. Especially, the Finnish archipelago is difficult due to varying grounds and winding fairways. The eastern part of the Gulf of Finland, nearby Russia, is an area where navigation situation is considered complicated due to minimal depths, large number of islands, varying climate, and narrow channels. (Deltamarin 2006, pp. 35−36.)

Cold climate and short summer time is limiting the boating season. Most of the year, especially during winter, the boats are kept in winter storage. The short boating season is not urging to invest lots of money into a boat. Therefore, many are satisfied with a smaller vessel that is easy to keep and maintain. Some of the vessels are rented. Boating trips are mainly domestic trips. The guest harbors are providing services, which seduce tourists to smaller retail shops. (Mäkilä 2009, pp. 8−9; Räsänen et al. 2005, p. 35.)

It is mandatory for a vessel that is equipped with an engine to have life jackets for every person on board according to the water transport regulation. However, it is not mandatory to wear them. A significant number of victims would have been saved if they had worn life jackets. Besides life jackets, the mandatory safety equipment on board include oars or a paddle, anchor provided with rope, emptying tool, emergency switch, tools for giving a distress signal, and an extinguisher depending on the size of the outboard engine. The most important safety factors are the common sense and judgment of the boater. The responsibility of the boat and people on board belongs to the owner or holder of the boat, or to the skipper during the trip. In addition, the vessel has to be safe. (Luntiala et al. 2004, p. 16; Trafi 2014)

Although most studies on recreational boating have investigated the rates of drownings, human errors and technical failures, the future discussion has a place to investigate under what kinds of conditions accidents happen. The statistics of boating accidents is not sufficient enough. In addition, statistics do not distinguish precisely in more detail, what kind of accident it is about. (Räsänen et al. 2005, pp. 46, 48.)

2.2 Maritime search and rescue related research on causes

Nowadays, the Gulf of Finland is an actively investigated part of the Baltic Sea. It is a narrow brackish water area and a complicated hydrographic region. In the Baltic Sea, the salinity is low, which makes the flora and fauna special. In addition to the significant environment, the Gulf of Finland is an important area for inhabitant, transportation and economic affairs. The area provides several options for recreational activities and trade opportunities, which are dependent on the sea area. Despite the popularity, the Baltic Sea, as well as the Gulf of Finland, is more vulnerable than many other sea areas. Maritime transportation and recreational activities have been increasing and are expected to increase in the future. (Viertola 2013, p.14.)

It is believed that the growing traffic is increasing the risk of accidents in the Gulf of Finland. In addition, the sea is freezing every year including significant variations on ice extent. Every harbor in the Gulf of Finland is covered with ice during harsh winters. The wind flows push the ice towards to the East, especially during mild winter times. Although the southern harbors remain open during mild winters, ships traveling to the ports located to the eastern Gulf of Finland are forced to pass through the stacked ice fields, which increase safety risk. The average appearance of ice in the Gulf of Finland is between December−April. (Kujala et al. 2009, p. 1349; Arola et al. 2007, pp. 2, 4.)

Although the ice and winter conditions are considered to be well-known, the behavior of the water masses under the ice is weakly known (Alenius et al. 1998, p. 120). The topography of the sea bottom is varying a lot. The depth of the Gulf of Finland is relatively low and the bottom is rocky. According to Kujala et al. (2009, p. 1351), the most common accident type in the Gulf of Finland during the years 1997−2006 was grounding. The next hardest group was ship-to-ship collision. The main cause of the accidents was a human factor. The majority of the vessels were cargo ships and tankers.

Recreational boating accidents caused by weather conditions have been studied relatively little in the world. Kokko (2013) has made a study on maritime search and rescue missions performed by the FBG in the Finnish sea waters. The accidents in the research were caused by weather conditions. Most cases were known to be caused by the following parameters: fog, wind, waves and icing conditions. The aim of the study was to recognize dangerous weather phenomena for boaters and to discover specific meteorological characteristics for dangerous weather conditions by using synoptic and mesoscale analysis. These were found quite well in the case of wind. Share of the fog turned out to be hard to estimate at that time and place where the mission happened. One of the reasons was that fog can be very local and thus hard to predict and observe.

The focus of the studies on boating accidents has mainly been on the risk factors such as behavior of the boaters, technical failures and human errors. O'Connor & O'Connor (2005) have investigated the causes and preventions of boating fatalities in Australia that happened during the years 1992−1998. According to the findings, nearly half of the vessels were lacking an appropriate number of floating devices for the people on board. A huge number of the victims were not using them. The number of deaths due to alcohol consumption was similar to road deaths. Nearly in a half of the fatalities alcohol has been involved in O'Connor and O'Connor's study. Drownings in Finland have been studied by Lunetta et al. (2004, p. 1053), who stated that more than half of the boat related drownings in Finland were associated with alcohol. The cases of the study happened from 1970 to 2000. Alcohol was a significant risk factor for drowning in Finland. The respective values for both males and women were more than 70 % on boating-related drownings.

The fatalities in O'Connor & O'Connor's study were mainly caused by human factors such as failure to wear proper devices. Also the inexperienced skills of the boaters were noted. Likewise O'Connor & O'Connor's conclusions about the causes, McKnight et al. (2007) stated that human error is the major cause of boating accidents in the U.S waters. Both studies address to invest in safety instructions for preventing hazards and address the proper use of safety devices. McKnight et al. (2007) also emphasized training skills for individual boat types, because errors led to boating accidents were varying among the vessel types. The popularity of different recreational vessel types has addressed the authorities to pay more attention on safety. Molberg et al. (1993) stated that vessels such as canoes, kayaks, rowboats, and inflatables were associated with a higher rate of fatal incidents than motorboats in Ohio 1983−1986. Also young age and lacking experience of the boaters were associated with the risk increase. In O'Connor & O'Connor's material a quarter of the vessels were recorded as dinghies.

The main causes of accidents are usually human failures, material or technical failures, or bad weather conditions which may lead to damages or losses of vessels or human beings, or personal injuries (Azofra et al. 2007, p. 942). Although the effects of weather conditions led to boating accidents have not been studied that comprehensively, weather phenomena are still believed to have an impact on boating accidents.

2.3 Maritime SAR system in Finland

Maritime Search and Rescue Act (1145/2001) involve assisting people in distress or danger at sea, also ships and vessels in difficult circumstances. It also provides first aid and radio communications. Besides these, Maritime SAR services provide ship and vessel assistance in a case of difficulty or emergency, accident prevention, search and rescue acts, and medical consultations and patient transportations. (Rajavartiolaitos 2014.)

Organizations that may attend maritime SAR operations in Finland are the following:

- the Finnish Border Guard (FBG),
- the Finnish Rescue and Emergency Services (FRS),
- the Finnish Police,
- the Finnish Transport Agency and Safety Authority (Trafi),
- the Finnish Defence Forces,
- the Finnish Meteorological Institute (FMI),
- social and healthcare services,
- customs,
- environmental authorities, and
- volunteers such as the Finnish Lifeboat Institution (FLI).

This study is focused on the missions operated by the FBG, the FRS, and the FLI. The FBG and the FRS are both official authorities. The FLI is a voluntary organization operating maritime SAR missions in Finland.

The FBG is the leading SAR authority in Finland. It is responsible for the coordination, planning development and supervising of SAR activities as well as coordinating cooperation with other authorities and volunteers. In a case of an emergency, the FBG is responsible for coordination of radio communications and enabling telemedical assistance services between the vessels and services that provide medical aid and care. The FBG is the head of the Maritime Assistance Services and responsible for receiving distress signals and delivering them to relevant authorities. It also provides SAR related training and education. (Rajavatolaitos 2014.)

Part of the missions in this analysis is from the FRS. In Finland, municipalities are responsible for the rescue cooperation in that area determined by the Council of State. Finland has twenty-two rescue areas, of which five are located at the coast of Gulf of Finland: Helsinki, Länsi-Uusimaa, Keski-Uusimaa, Itä-Uusimaa, and Kymenlaakso. For rescue acts every area has a rescue department. (Sisäministeriö/Pelastusosasto 2014.)

The Ministry of Interior is the leading body of rescue services. Rescue operations and other task related to rescue or emergency are carried out by regional rescue departments, which are responsible for the maintenance of the rescue service system and provision of the services. Rescue departments must also distribute rescue service related knowledge and guidance, provide expert consultations, perform task that are expected of rescue authorities (accident prevention, damage limitation and fire inspections), perform their part in civil defense acts and maintain the state of preparedness required by them. Rescue departments also have to train their personnel and act as intermediary to fit together operations of other authorities and parties that attend or provide rescue services. The Rescue Departments also carries out oil spill prevention and response operations and any other tasks that are addressed to the rescue acts. (City of Helsinki 2013; Sisäministeriö/Pelastusosasto 2014.)

Voluntary organizations may also attend to search and rescue activities according to their own regulations and rules, which have to be approved by an official authority. The FLI is a leading voluntary organization for maritime rescue associations in Finland. The aim of the organization is to promote safety and good seamanship in boating. The rescue operations are based on voluntary participants. The major group of recipients is formed by recreational boaters. The cases are the most commonly caused by technical factors or unexperienced knowledge of seamanship. The FLI is a member of the International Lifeboat Federation (ILF). (Suomen Meripelastusseura 2014.)

2.3 Risk-informed decision making for SAR location planning

A sea rescue operation can technically be considered as an activity aiming at rescuing persons, vessels or any items from immediate and unavoidable danger which requires action. The main task of maritime rescue services is to prevent accidents and execute missions that occur in the marine environment. A sea accident is an event which has an impact on the vessel, the crew on board, the cargo or the environment. However, the concept of sea rescue has changed during the last decades and so is the so called traditional sea-rescue model. In the case of an accident or an emergency, the data is compiled by the authority in charge. This accident data and location information is used for planning of distribution of sea rescue resources. (Azofra et al. 2007, pp. 941–942.)

Azofra et al. (2007) introduced a gravitational model for assessment of location evaluation taking several risk factors into account. The model is aimed to relocate or distribute the adequate rescue equipment efficiently. The model analyses each resource individually. The addressed factors consist of characteristics of accident, vessel and damage, type of the accident, level of severity, distribution of rescue resources (meaning equipment such as helicopters, tug-boats and rescue vessels including their radius of action), placement of resources, and cost-effectiveness. According to Azofra et al. (2007, p. 949), wide range of technical factors is taken into account when assigning rescue resources. By using the factors helps us to assign weights for the suitability of a given location. The presented model was developed in the means of political decentralization and can thus make so called objective decisions about the locations of rescue resources. It was claimed in the study that political factors have the major impact on locating rescue resources.

Response time is critical in rescue operations. The response time depends on technical properties, such as speed and operability, distance and weather or environmental conditions, length of the coast, traffic types, traffic volumes, danger types in traffic, accident rates, and adequate facilities such as seaports, airports, hospitals and other rescue departments. The suitability factor of a location for rescue resource depends on many aspects such as the access to the location in various sea and wind conditions, and the availability of appropriate port infrastructure and services. (Azofra et al. 2007, p. 944.)

Besides response time, distance is also a significant parameter. It is the only controllable parameter once the rescue resource is determined (Azofra et al. 2007, p. 944).
Optimization models, introduced by Li (2006), were applied in search and rescue for the location of search and rescue units. The main goal of Li's study was to modify and apply several linear programming models in optimizing locations for SRUs. Also a simulation model was employed. The simulation analysis turned out to be important to calculate the significant statistics such as coverage, response time, and workload balance.

Recently, the interest has been focusing on such questions as which types of SRUs would be suitable for the operations in the Gulf of Finland. Goerlandt et al. (2013) introduced advances of a simulation model, which is aimed to evaluate the response characteristics of voluntary maritime SAR system. The aim of the model is to provide elements that support decision making to the FLI. The model is expected to give information regarding the adequate number of vessels in each rescue station and which type of SRUs are preferred to invest in by constructing a simplified representation of the performance of the SAR system. According to Goerlandt et al. (2013, pp. 1, 4), the performance is mainly driven by response time, which is dependent on the achievable speed of the SAR vessel in the prevalent wave conditions. Additionally, the wave conditions are dependent on location and time. Also fleet reliability was considered as a significant factor in the response capability of the SRU types. The model may help plan to acquire proper material for the SAR service.

Goerlandt et al. (2013) points out that although the historical data is very relevant for SRU location planning, there are several other factors that have to be taken into account. The authorities in charge may not measure such data concerning environmental and several other factors such as distances, vessel speeds or wave conditions. Norrington et al. (2008) have applied Bayesian Belief Network (BBN) to model the reliability of SAR operations in the UK Coastguard Coordination Centers. The BBN presents the probabilities of the success of SAR operations taking all relevant variables into account as well as important factors that are believed to have an impact on the performance of the SAR operation. The variables used in the model were favorable weather, competency of resources (SAR personnel), acceptability of response time, number of concurrent incidents, severity of incident, known location, effective communications, competency of MRCC (incident coordination) staff, and casualty situation. The main finding of the study was that the construction of a BBN can be complemented by using statistical analysis of secondary data. This provided more objective picture of relationships between the variables. However, it was restricted in the level of detail due to the lack of available data.

According to these studies, the relevant parameters for SRU location planning as well as for the adequate equipment proposal are the response time (including start time, travel time, operation time and end time), location, distance, severity of incident, number of incidents, and distribution of rescue resources. These factors are, however, influenced by several other factors.

3 Visual data mining of multidimensional data

The purpose of visual data mining is to enable humans to get an insight and make conclusions of the data by presenting the data in a visual form. This comes in use if the data is unfamiliar and the goals of the exploration are not clearly defined. Techniques for visual data mining, according to Keim (2002, p. 100), have a high capability for exploring data of voluminous databases and are proven to be highly valued for explorative data analysis. In addition, there are no requirements for the user to understand complex mathematical or statistical methods due to automated interactive tools. This allows the user to get a direct involvement with the data and to come up with new hypotheses.

A large dataset, in this context, means that we have a large number of objects. High-dimensionality, correspondingly, means that the dataset has a large number of variables. Geographic data tends to be high-dimensional and may be compiled from several different data sources. The data, therefore, often includes information that is irrelevant. But for finding the relevant attributes, specific patterns and combinations of relationships, the large number of variables is needed in order to find out, which variables are interrelated. (Guo et al. 2005, p. 115).

The subsections of this chapter deal with the visual exploration process, challenges and the techniques used in the analysis. The characteristics of multidimensional data are first discussed in section 3.2.

3.1 Visual data mining and exploration

Visual Analytics can be defined as a combination of Model Visualization and Exploratory Data Analysis. Model Visualization is used for making the discovered results understandable and interpretable by a human. Exploratory Data Analysis is a human-guided approach and is known as interactive exploration using graphical representations that support data mining tasks. (Ferreira de Oliveira & Levkowitz 2003, p. 379.)

The process in visual analytics usually combines automation to visual analysis methods providing human interaction. Data mining methods are commonly automated and used for creating models out of the data. This requires data interaction in order to hone the data. Visualization of automated methods enables user to make selections, modify parameters or use tools. Model visualization is for evaluating the results. (Keim et al. 2010, p. 10.)

Creating effective visual representations is a laborious task and requires many steps, which often tends to be a loop (Figure 4). It requires understanding of the visualization pipeline, the characteristics of the data, and the analytical tasks and techniques that are at hand. Most software used in visualization is developed with incomplete information about the available data that are used in the task. Therefore, there is a need for the data pre-processing. It eases the process if some of the different interactive operations are grouped together. The main thing, however, is to think about the users' intentions and purpose of exploring the data. (Thomas & Cook 2006, p. 11; Dix et al. 2010. p. 119.)

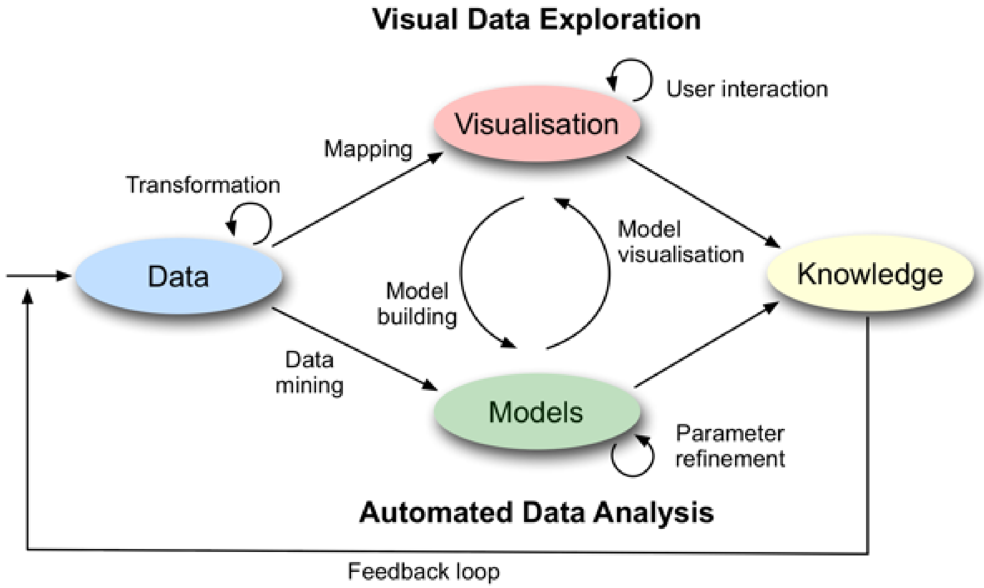


Figure 4 *Visual data-exploration is a loop that requires human interaction. (Keim et al. 2010, p.10).*

3.2 Multidimensional data

One-dimensional data may be related to temporal data, for example, time-series and trajectories. One-dimensional data usually does not have location information. Two-dimensional data may be a geographical map representation where the two dimensions would be longitude and latitude in an x-y-plot. If the data consists of more than two or three attributes, the data is known to be multidimensional or multivariate. It usually requires more sophisticated visualization techniques rather than two- or three-dimensional plots. Geographic datasets are often multidimensional. Multidimensional data consists of hierarchies of different levels with many variables. The data includes, besides geographical location information, also time period and characteristics and/or responses of the variables. One example would be studying the relationships between the variables over a time period. (Keim et al. 2008, p. 164.)

The relevant aspects of multidimensional data are space and time, and their relationship. Uncertainty has an impact on the data quality. This section introduces the characteristics of spatial and temporal data and briefly reviews the uncertainty of the data. The source of this section is from a chapter of Keim et al. (2010, pp. 57−86) written by Andrienko et al. (2010). The chapter handles the space, time, and uncertainty of data. The information of the chapter (Keim et al. 2010, pp. 57−86) is reviewed in this section.

3.2.1 Spatial and temporal dependence

Spatio-temporal data is both constrained and upheld by the concept of spatial and temporal dependence. The spatial domain refers to the Tobler's first law when "everything is related to everything else, but near things are more related than distant things." This means that features near each other tend to be correlated, which is called spatial autocorrelation. (Andrienko et al. 2010, p. 63.)

This law, however, is not determinate. Geographical space is not homogenous, which weakens the law. Every location has some level of uniqueness that is related to the location. This is known as spatial dependence, which can be affected by natural or artificial barriers, such as climate or a seasonal change. (Andrienko et al. 2010, p. 63.)

Besides the distance, also direction may define a dependency between the variables. Closeness is also dependent on the phenomenon, which may be spatially defined, for example, in terms of distance by roads, rather than straight line distance. Temporal dependence may be affected by events, for example, changes caused by natural catastrophes. Events in time have an effect on the future rather than past events. So called temporal distances may be measured in terms of working days, number of hours or past years. (Andrienko et al. 2010, p. 63−64.)

3.2.2 Characteristics of spatial and temporal data

Time dimension may consist of single or multiple levels of scale. Temporal primitives can be aggregated or disaggregated into larger or smaller units, for example, 60 seconds into one minute or vice versa. In spatial analysis, the scale may affect the results. For example, some patterns may be detected at one scale, but they may not at another scale. In an extreme case, opposite relations may occur in the same place or time in different scales. This behavior is known as scale dependent. (Andrienko et al. 2010, p. 65.)

The scale of the analysis must match with the scale of the phenomena under consideration. This turns out to be a key problem in analysis. It is not an easy task to understand which scale would be appropriate. Data units can be aggregated into larger units various ways. Disaggregating the size, however, is only possible with additional data involved. The results of the analysis may be dependent on how the units are aggregated. This refers not only to the size of the aggregates but also to the location and composition of smaller units. Testing the sensitivity of the findings to the means of aggregation is necessary. (Andrienko et al. 2010, pp. 65–66.)

When choosing the scale, the goal of the analysis has to be known. Various scales of geographic and/or spatial phenomena interact and some phenomena may occur from smaller or larger phenomena. This is the notion of hierarchy of scales. (Andrienko et al. 2010, p. 67.)

Time has a hierarchical system of granularities consisting of seconds, minutes, hours, days, weeks, months, years, centuries, etc. These types of granularities are organized differently in different calendar systems. Time contains cycles and re-occurrences. Some of these might be regular and easily predictable (seasons), others are less regular (holidays, social cycles). The temporal dimensions can be considered as time points (an instant in time) or intervals (a temporal primitive with an extent). Selecting the appropriate primitives depends on the data properties and the problem under consideration. Transformations are often required when dealing and processing the temporal data. (Andrienko et al. 2010, p. 67.)

Different temporal structures have been defined for time: ordered time, branching time, and multiple perspectives. Ordered time can be divided into two sub-classes: linear and cyclic time. Linear time is often considered as a continuous sequence of temporal primitives. It proceeds from the past to the future. Cyclic time span is a composition of a finite set of recurring temporal primitives, for example, the time of day or the seasons of the year. The concept of branching time involves with the description and comparison of different scenarios, which is commonly used in planning. Multiple perspectives enable several points of views for the facts under consideration. This perspective is used when, for example, people are sharing observations about events via media. (Andrienko et al. 2010, p. 67.)

3.2.3 Uncertainty of data

The quality of the data is often affected by errors, missing values, deviation of other sources, which weakens the quality. The reason may lie in several factors, for example, in data acquisition, transmission, or processing the data, which may result in information loss. Uncertainty can be treated as a combination of different aspects of uncertainty, such as error (e.g. outlier or deviation), imprecision (e.g. resolution of a value), accuracy (e.g. size or scale of interval), lineage (e.g. data source), subjectivity (e.g. subjective influence), non-specificity (e.g. lack of distinctions for objects), or noise (e.g. background influence). Geometric uncertainties may be found in geospatial, time and thematic data uncertainties, which might be quite different concepts from each other. The distinction between the uncertainties might therefore require special treatment. (Andrienko et al. 2010, p. 64.)

3.3 Steps dealing with data

Before techniques can be applied, the multidimensional data need to be transformed and processed in a way that the application is able to handle them according to the purpose of the task. Common preprocessing tasks are, for example, data cleaning, transformation, selection, reduction, and integration. (Keim et al. 2010, p. 10.)

Data cleaning: Data cleaning can be used for removing missing values and correcting inconsistencies in the data. There are several channels that turn the data into an inaccurate form. The flaws may occur due to a poorly designed input, human error, issues in delivery or data share, the data has gotten outdated or wrong methods or procedures have been used. (Han & Kamber 2006, p. 65.)

Data transformation: In data transformation the data are consolidated into appropriate forms for the task. This step may include smoothing (binning, regression, clustering), aggregation (summarizing, merging), generalization (low-level concepts into high-level concepts in hierarchies), normalization (data scaling, min.–max.), and attribute construction (new attributes are created and included to the given set of attributes). (Han & Kamber 2006, pp. 70–71.)

Data selection: Selecting data aims to gather all the appropriate data, which are the best by the quality. Compiled data includes the widest data range for the problem solving. Data selection may be performed by using tools (e.g. brush and highlight) and classification methods according to certain criteria.

Data reduction: The data derived from multiple sources consists of several types of different data that are more or less related or not related to one another. The task of data reduction is to reduce the data amount to ease working with the data. Concept of hierarchies as well as data generalization and aggregation can be used as a form of data reduction (Han & Kamber 2006, p. 47).

Data integration: Data integration combines data from multiple sources into one store. The data may derive from different databases, files or other systems that may use other formats (Han & Kamber 2006, p. 67.)

3.4 Categories of visualization

Visualization techniques are useful for giving an overview of the data. For analyzing the interesting patterns and subsets of the data, the user needs to get closer to the details of the data. In this step, it is important to keep up the overview step while focusing on the target group and using other visualizations on other targets.

There are a large number of techniques for visual data exploration and data visualization. The techniques can be applied to all steps of the analysis process. Keim (2002, p. 1) has classified the visualization techniques based on three criteria: the data to be visualized, the visualization technique, and the interaction and distortion technique used.

In this work, the relevant properties of 'Data to be Visualized' are one-, two- and multidimensional data. From the 'Visualization Technique' the relevant topics for visualization in this work are somewhat 'Geometrically-transformed Display', and 'Standard 2D/3D Display', because the multi-linked views are used.

Section 3.6 lists the relevant and the most common interaction techniques.

As mentioned in section 3.2, one-, two-, and multidimensional data can be visualized. *Hierarchies, graphs, algorithms* and *software* are often visualized also. Graphs are commonly used for describing relationships between other classes or variables. A graph may consist of items (nodes) and connections (edges) that bound these objects together based on the characteristics of the items. One of the well-known methods for visualizing hierarchies is tree visualizations. Graphical techniques are used in software visualization tools to make software visual with program displays, artifacts, and behavior. Program code languages are visualized to show hierarchical data. (Keim et al. 2008, p. 164; Ball & Eick 1996, pp. 2, 7.)

Geometrically-Transformed Displays: The goal of geometrically transformed displays is to find interesting transformations of multidimensional datasets. This includes techniques from exploratory statistics that are commonly used for data processing. For example scatterplot is a traditional technique, where two data variables are projected on the *x* and *y* axes of a Cartesian coordinate system. (Ferreira de Oliveira & Levkowitz 2003, p. 380.)

Other visualization techniques that Keim (2002) has presented are *iconic displays, dense pixel oriented displays* and *stacked displays*. The idea of iconic displays is to illustrate the multivariate data feature with an icon or a glyph, which can be arbitrarily defined. They can be, for example, faces, star icons, whiskers, stick figures, color icons, and tile bars. (Ferreira de Oliveira & Levkowitz 2003, p. 380.)

The idea of *dense pixel-oriented display* techniques is to map data values to colored pixel and group the pixels into adjacent areas or subwindow according to the dimension the values belong to. Such techniques are, for example, space-filling curves, recursive patterns, and circle segments. Stacked display techniques are presenting data by partitioning it and building hierarchies. Example methods may be treemap, cone trees and dimensional stacking where one coordinate system is embedded inside another coordinate system. (Keim et al. 2003, pp. 41−42; Keim 2002, p. 104.)

3.5 Visual variables

Visual variables, described by Bertin (1967|1983), are often used for mapping spatial data (Figure 5). Besides these, there is one additional variable that is used for describing quality–clarity. Clarity can be divided into three subdivisions: crispness, resolution, and transparency. Crispness refers to the visible detail of map elements such as edges or fill, or their combination. Resolution deals with spatial precision change. In a digital environment, the size of the grid has an effect on the outcome of the display, for example, which raster cells or vector data will be plotted. Resolution defines the level of precision in the spatial data underlying an attribute. Transparency is considered as the degree in which a theme can be seen through a "fog" that is placed over the theme. Transparency is often used with the color value, another visual variable, to illustrate uncertainty. It can be seen out of the behavior how the magnitude of transparency changes with the fuzziness over the theme. (MacEarchen 2004, p. 276; Yao & Jiang 2005, p. 222.)

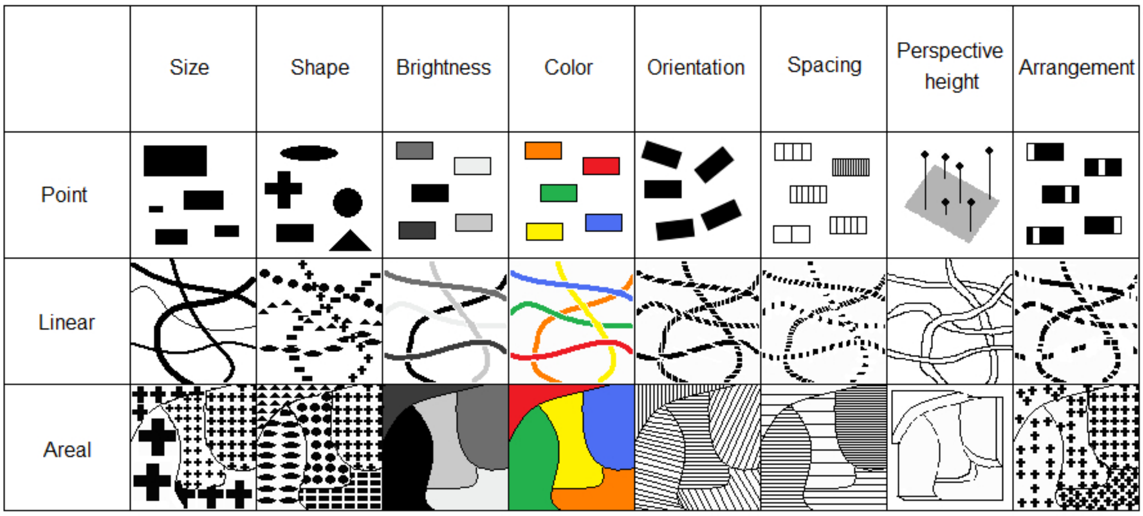


Figure 5 *Visual variables used for spatial information. (Ward et al. 2010, p. 216).*

In this study the most relevant visual variable is the color hue. Brightness and saturation, which is not listed in Figure 5, are also involved with the color. These two variables are briefly described in section 3.7.2.2. Transparency is related to the PCP technique.

3.6 Interaction

The importance of interaction in visual exploration is being emphasized, because it provides the user with the opportunity to explore the dataset. Thomas and Cook (2005, p. 75) present the three ways to look at the science of interaction: human time constants, use of interaction, and nature of interaction. The first viewpoint is important because all means of interaction are constrained. This defines what the user is capable cognitively and perceptually. For example, for producing the perception of an animation the necessary rate for a time constant is around 100 milliseconds, in order to produce the perception of an immediate response. The second viewpoint deals with the use of interaction to execute and accomplish tasks such as data manipulation, manipulation of mappings, navigation, and dialogue. The third viewpoint deals with the use of interaction itself, including the difference between the interactions in their environment (2D and 3D) and the effects of the devices used for the interaction.

The basic categories of interaction are the following (Dix et al. 2010, p. 119):

- select: mark data items that are interesting,
- explore: show some other data by using, e.g., zooming or panning,
- reconfigure: rearrange the data,
- encode: change visual appearance, e.g. color, size or shape,
- abstract/elaborate: increase or decrease details, e.g. aggregate,
- filter: select data that matches with certain criteria,
- connect: highlight related data, e.g. selection by brushing shown in multi-linked views.

3.6 Challenges in data visualization

Highly developed systems automatically restore data in large repositories without filtering and refining the data into a usable form. Datasets may include information that may be useless for the purpose of the task, transformed or processed inappropriately or presented in an inappropriate form or way. This leads to the danger that the information is getting lost in data or datasets. This is known as the *information overload problem*. (Keim et al. 2008, p. 154.)

Voluminous datasets are built by information that derives from different data sources, which makes the structure of the data heterogeneous. Compiling huge amounts of data without any associated metadata (such as what it is, where it comes from, from where it was taken, how, etc.) restricts the opportunity to deal and interact with the data and gain more insights from it. The main problem among the huge data pile is to find suitable methods and models, which can transform the data into an ideal form that is legible and supports knowledge improvement and, correspondingly, decision making (Stikeleather 2013; Keim et al. 2008, p. 155.)

Although, we have automated technology and powerful tools for analysis it is not always understood or analyzed properly. The available tools for a particular analysis task may require specific equipment or user skills in order to work reliably for problems that are clearly defined and understood. This makes the decision making process even harder. Secondly, the communication in sharing information of automated data between the publishers is lacking. The elements of the visualization may be processed by, for example, color coding, positioning, and other visual effects, which may have an influence on the data interpretation. This may end up for specifying patterns and outcomes beforehand with algorithms that are used by the professionals. This may have a significant impact on the visualization and its interpretation. Not everybody may understand it. (Keim et al. 2008 p. 155; Stikeleather 2013.)

As the visualization approach provides a direct contact to the data characteristics by alternative displays and tools, the techniques may not be effective enough due to large datasets that may be increasingly common. The visualization might turn out to be illegible due to cluttering and overplotting with a large number of records. In addition, the technology may not be sufficient enough for updating the display or responding fast enough to user's interactions. This makes perceiving, tracking and comprehending dynamic data challenging for the user. (Keim et al., 2010 p. 77.)

3.7 Techniques for visualizing and exploring data

3.7.1 Parallel coordinate plot

Parallel coordinate technique, developed by Inselberg and Dimsdale (1990), is one of the most common techniques in discovering dependencies between the variables visually. It was first developed for n-dimensional geometric computations and succeeded in development of complex algorithms and in optimization problem solving. (Siirtola & Räihä 2006, p. 1279.)

The advantage of parallel coordinates is the opportunities in handling and dealing the multivariate dimensions both in a general and detailed manner. Secondly, it allows user to interact with the data in several ways. Siirtola and Räihä (2006) are demonstrating the importance of the interaction in visualization by presenting the interaction techniques used in parallel coordinates and comparing them to the guidelines of visualization designs in literature.

It is said that gaining an overview of the data set with the parallel coordinates is very simple, because the plot itself is already an abstract overview. The characteristics, such as clusters, outliers and missing values, can be explored as well as the relationships between the items as in the Figure 6 it is shown in a general overview. The overview includes all the used values and items, but still only a subset or parts of potential relationships between the variables. This must be taken into account, because significant connections might get unnoticed if the corresponding axes are not next to each other. (Siirtola & Räihä 2006, p. 1286.)

The comparison of dependencies between the variables and the patterns depends on the order of the axes. Rearrangement of the axes is important because the connection appear only for adjacent axes. The aim of the rearrangement is to discover interesting relationships between interesting attributes and to find new ones that was not assumed before. (Chen & MacEarchen 2008, p. 262; Siirtola & Räihä 2006, p. 1294.)

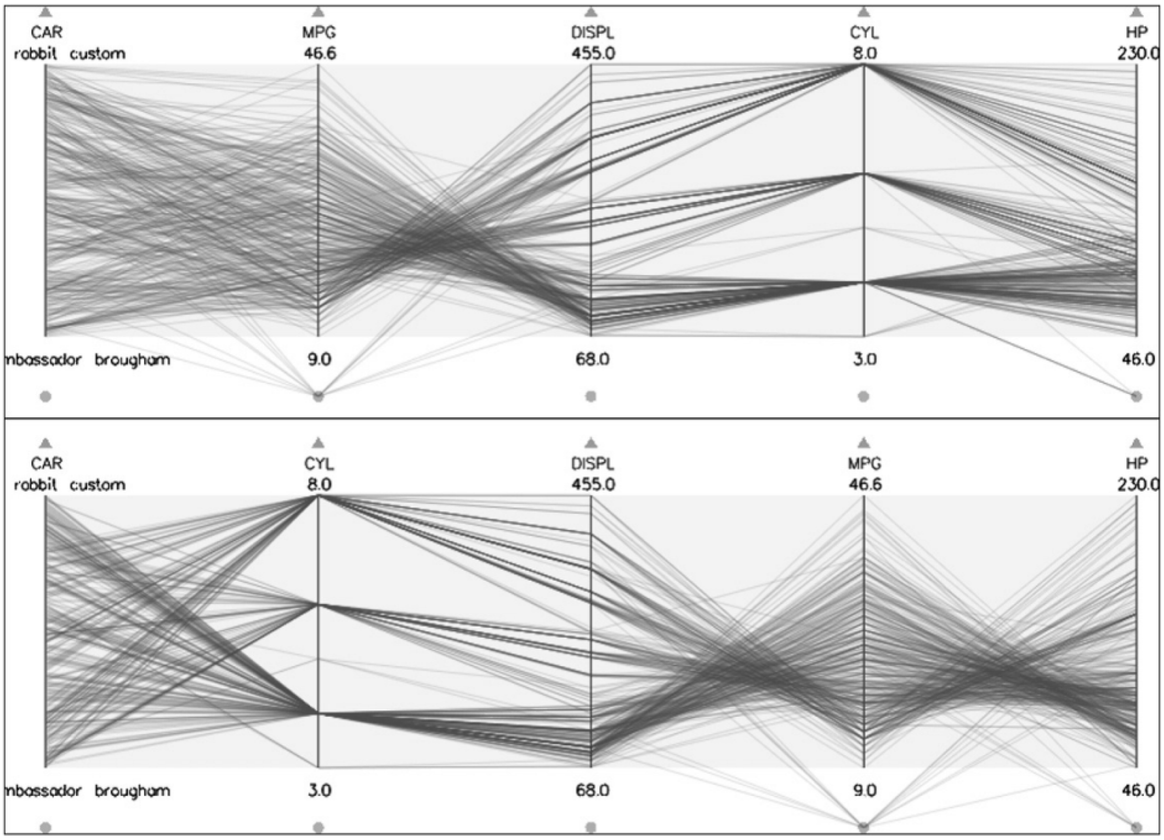


Figure 6 *An overall view of the data plotted in the PCP. (Siirtola & Räihä 2006, p. 1288).*

Parallel coordinate plot (PCP) maps multidimensional dataset onto a two-dimensional display of a set of axes. Every group of attributes in the dataset is defined as a vertical axis in the parallel coordinate plot (axis may also be drawn horizontally). They are placed vertically and parallel next to each other. A single data item is depicted as a polyline that are placed horizontally (may also vertically) connecting the vertical axes.

Point and line duality is a typical feature for the parallel coordinates. A data item is represented on a two-dimensional as a connecting line of the axis pair. If the group of points is intersecting between the axes on the plot, the corresponding points are lying on the same line in an X and Y coordinate plot. Each data item is presented as a polyline between axes. The connections are either parallel or they intersect in the same point. It is possible to detect which polylines are parallel at the same point and which are intersecting the point. In a high-dimensioned data we must check the set of intersection point, which may appear between the parallel axes, on the axis, or outside a pair of axes. (Siirtola & Räihä 2006, p. 1284.)

The parallel coordinates aim to make the correlations between the data visible. Correlation is the relationship between the variables. There are three different correleations: a positive correlation, a negative correlation, and a zero correlation (or no correlation). (Peterson 2009, p. 8.)

The different types of correlation are presented in Figure 7. A positive correlation would make the polylines go parallel horizontally (or increasing slope of points). There is a negative correlations if the lines are intersecting at a point (or a decreasing slope of points). Zero correlation would set the points vertically or horizontally.

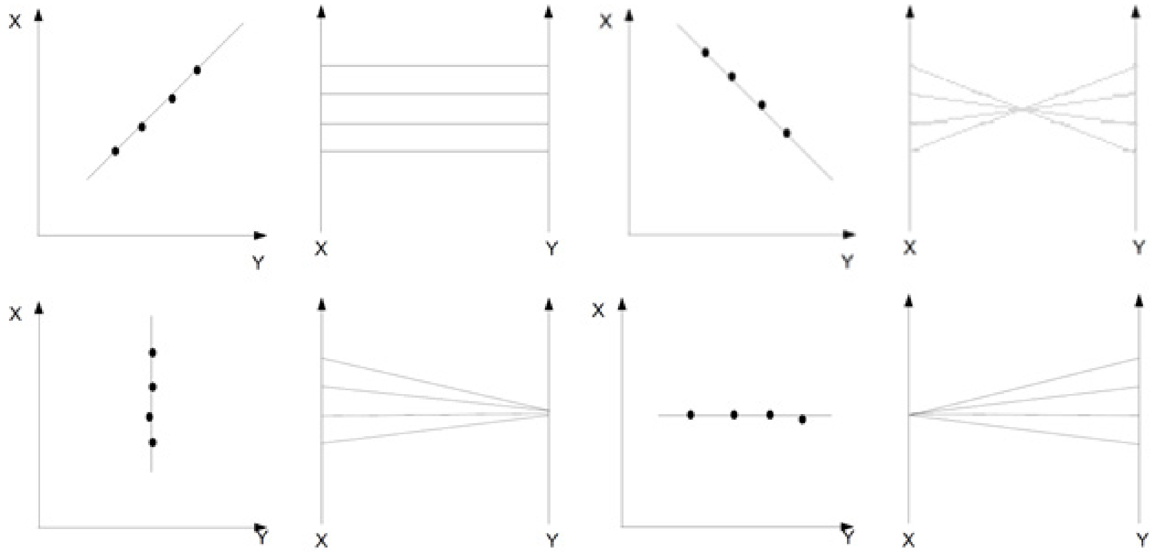


***Figure** 7 Correlations represented in PCP: positive correlation (upper-left window), negative correlation (upper-right window), and zero correlation (lower-left and –right windows).*

The PCP allows the user to visualize datasets with high number of attributes. It is possible to add additional variables and polylines to the PC plot. The number of adding polylines and variables depends on the size of the display and the limitation of the visualization by the human eye. Although, a PCP is suitable for detecting multidimensional data in detail, overplotting may cause problems in visualization even with a decent amount of records. (Peterson 2009, p. 7; Chen & MacEarchen 2008, p. 262.)

In the case when there are a massive set of polylines, the visualization will not bring much detail to the plot, because the colors of the polylines do not distinguish from each other that clearly. The colors of the polyline create an even surface. This is called as *occlusion problem*. This causes a big data loss because the overlapping makes it impossible to differentiate the polylines from each other. (Siirtola & Räihä 2006, pp. 1294–1295.)

Other common problems for parallel coordinates are *ambiguity* and *clutter* (Figure 8). In the case of ambiguity problem, the plot is identical for two sets of data points. In Figure 9, the identical sets of points will be (1, 2, 1) and (2, 2, 1) as well as (1, 2, 3) and (2, 2, 3). Clutter arises from the number of items and ambiguity is occurred by the number of attributes or axes (Peterson 2009, p. 10).

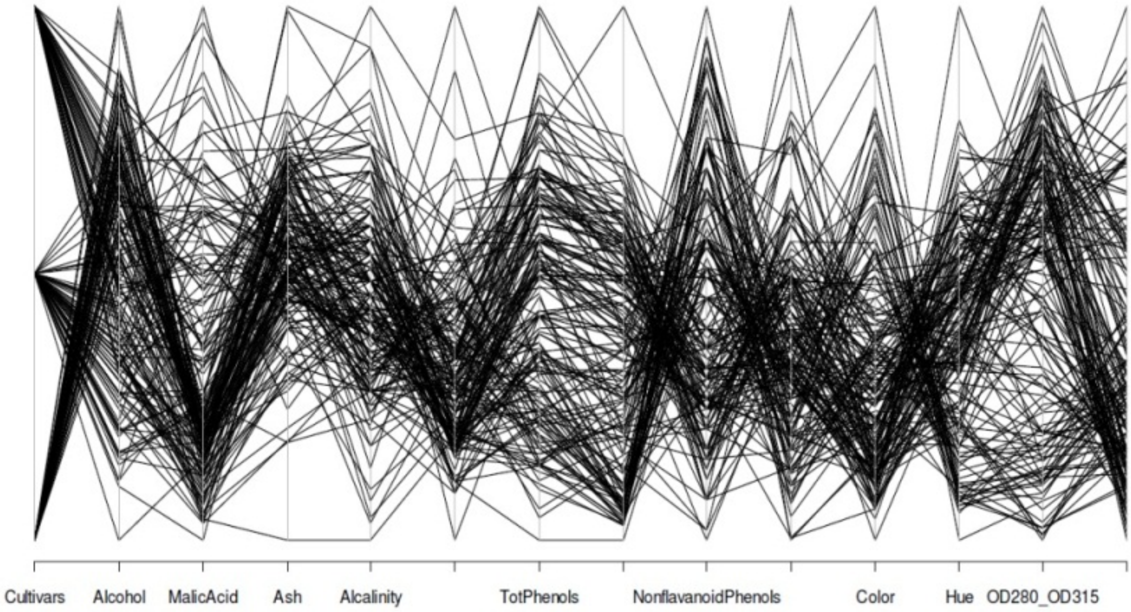


Figure 8 *Ambiguity and clutter on the same plot. (Peterson 2009, p. 15).*

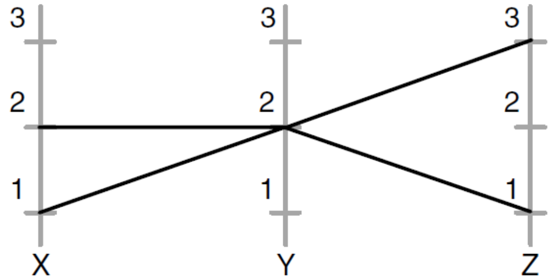


Figure 9*(edited) Identical sets of points (1, 2, 1), (2, 2, 1), (1, 2, 3) and (2, 2, 3). (Siirtola & Räihä 2006, p. 1283).*

One solution for these problems is to visualize data abstraction, such as groups and clusters (Chen & MacEarchen 2008, p. 262). Ambiguity and occlusion problem may be solved by brushing and highlighting a polyline or a group of polylines (Figure 10). With the brushing operator the selected items get highlighted, which makes it easier to apply other operators, such as zooming into them and getting more detailed information out of the data. This enhances the data comparison as well as data reduction. This avoids the ambiguity problem by selecting a polyline which can either be compared to the whole data set or to the current selected other items. (Siirtola & Räihä 2006, pp. 1289–1290.)

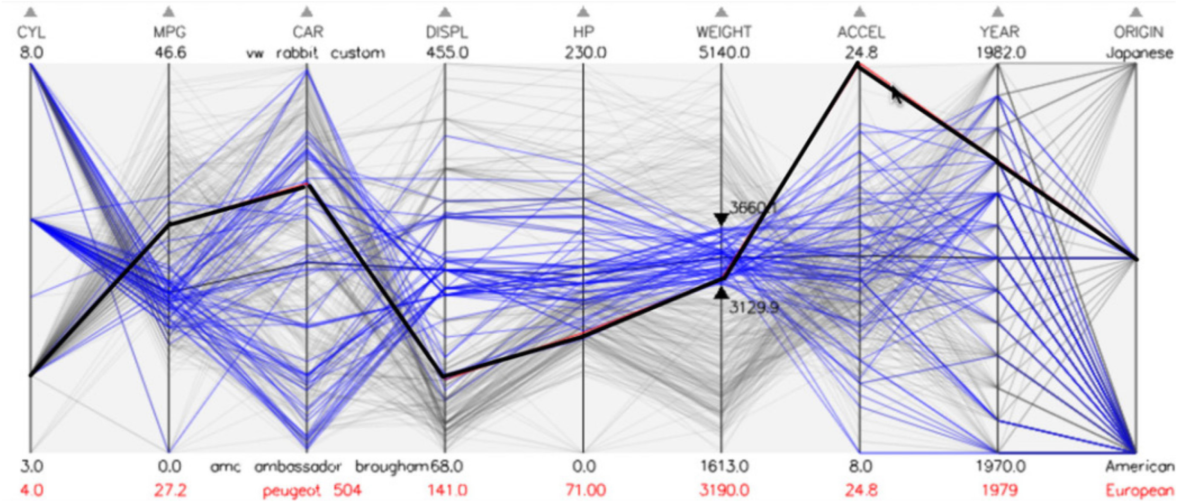


Figure 10*(edited) Brushing of a single polyline. (Siirtola & Räihä 2006, p. 1293.)*

The other option for the solution of the occlusion problem is making the polylines transparent. This makes the clustered polylines look darker and thus easy for selection. The abstraction can also be processed by summarizing the data by observing, for example, the arithmetic mean and standard deviation with a thick line and boxes that are placed on each axis. The size of the boxes is defined after the standard deviation of the values. The thick line is indicating the arithmetic mean of the values which goes through the axes. The method is shown in Figure 11. (Siirtola & Räihä 2006, p. 1295.)

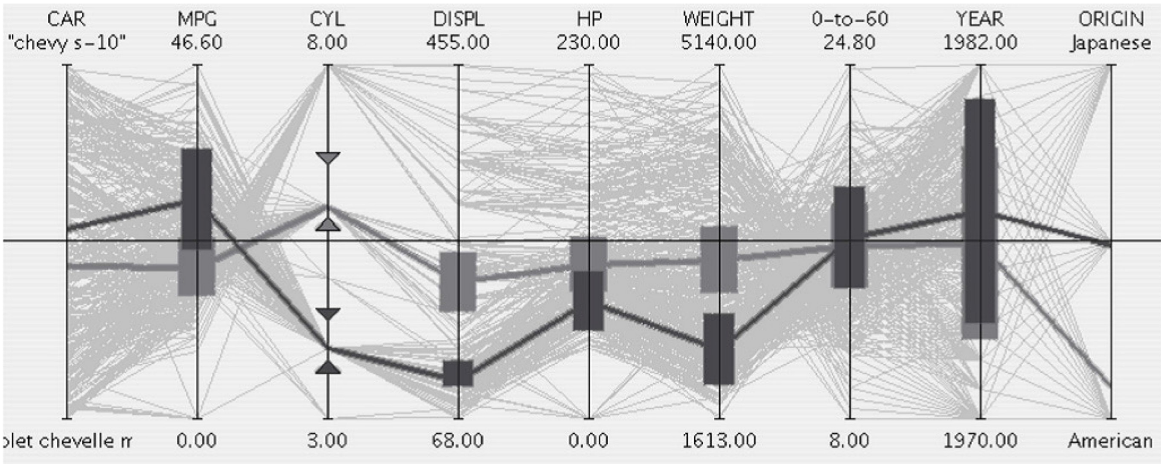


Figure 11*(edited) Thick line passing through the axis is representing the arithmetic mean. The size of the boxes is defined after the standard deviation of the data values. (Siirtola & Räihä 2006, p. 1296.)*

3.7.2 Bivariate map

3.7.2.1 Theory of bivariate map

Thematic maps usually show one data attribute such as income per capita. This kind of map presentation is a univariate map. More interesting and informative thematic maps use juxtaposition of two or more attributes. This way it is easy to compare different attributes and help us recognize important connections. (Axis Maps LLC 2009–2010.)

In the field of Cartography, bivariate mapping is an important technique. It maps two variables of the set of geographic features on a single map. Those two variables are combined with two different sets of graphic symbols or colors. The functional purpose of a bivariate map is to represent relationships between two variables. Displaying two or multiple attributes on a map is known to be an efficient way to observe statistic information. However, designing tends to be challenging due to the high level of density of the information. (Liu & Du 2014; Elmer 2012.)

For example, a bivariate map that represents the income per capita and life expectancy rate might unveil a strong correlation between the variables. This can be indicated with two separate colors, one color representing one attribute (Figure 12). Multivariate maps also help to save space. More data can be added on a single map rather than spread them across a series of maps of single data. (Axis Maps LLC 2009–2010.)

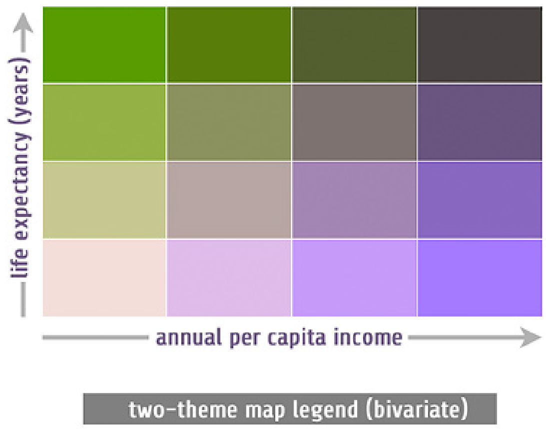


Figure 12 *Color scheme depicting correlation between two attributes. One color hue represents one attribute. (Axis Maps LLC 2009–2010). Source: http://indiemapper.com/app/images/bivariateLegend.jpg. [Referred 21.1.2014].*

Visualizing bivariate data on a map is more complex compared to univariate data. The complexity makes the mental process for the viewer more difficult. If this issue turns out to be intractable to comprehend, the utility of the map decreases for the viewer. Maps representing two or multiple attributes easily become cluttered and overlapped with symbols and colors. For example, a bivariate choropleth map often requires the user to return to the map legend in order to read the map. This is why the colors are kept straight to help to read the map. Multivariate maps are rich by their content and thus require effort to understand the data. (Elmer 2012; Axis Maps LLC 2009–2010.)

3.7.2.2 Bivariate color sequences

For bivariate technique, the most relevant visual variables are color, brightness (also lightness) and saturation. Other variables such as the dynamic variables are briefly taken into account in the analysis. Color can be defined by the two parameters, hue and saturation. Hue is kept as the actual consideration of color, the dominant wavelength. Brightness indicates how dark or light the color is. One easy example is showing the brightness or darkness as shades of gray (Figure 13). Brightness can be used to depict relative importance, order, or numerically classified data. General guideline for using brightness is ‘the more the darker’ meaning if the value higher, the darker the color. (Slocum et al. 2009, p. 84; Brown & Feringa 2003, p. 134.)

***Figure 13** Shades of gray indicating brightness. (Ward et al. 2010. p. 142.)*

Saturation is the purity of the color, the level of hue relative to gray (Figure 14). Color use in mapping is usually used map data values so that the relationship between the value ranges and color values are defined. Continuous range of hue and saturation values is used in color maps or figures usually for interval and continuous data variables (Figure 15). Colors are also generally selected for individual data values to show the distinction between data objects. (Ward et al. 2010, pp. 143−144.)

***Figure 14** Blue and yellow colors related to gray. (Ward et al. 2010, p. 143).*

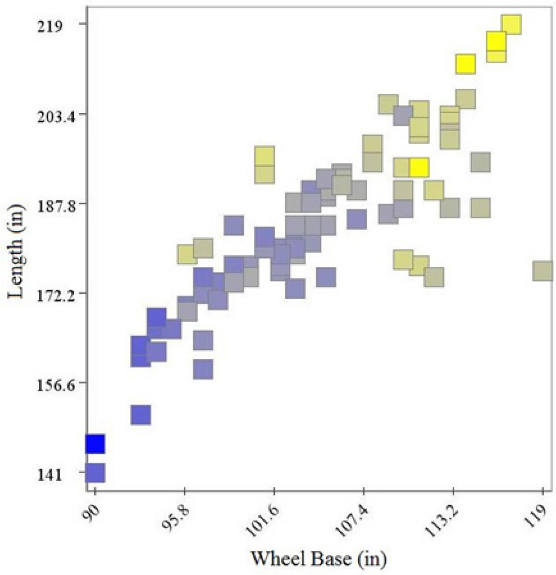


***Figure 15** Saturated colors used in data values. (Ward et al. 2010, p. 143).*

Color is considered as three dimensional. Displaying two or three dimensions is possible by using pseudocoloring, which is a method of assigning arbitrary colors to the levels of gray. The image is usually black-and-white, for instance, a satellite image. On a map, it is preferable to map data dimensions to perceptual color dimensions such as hue (variable 1) – saturation (variable 2) or hue (variable 1) – lightness (variable 2). One example map is presented in Figure 16, which uses bivariate color sequence. (Ware 2013, p. 134.)

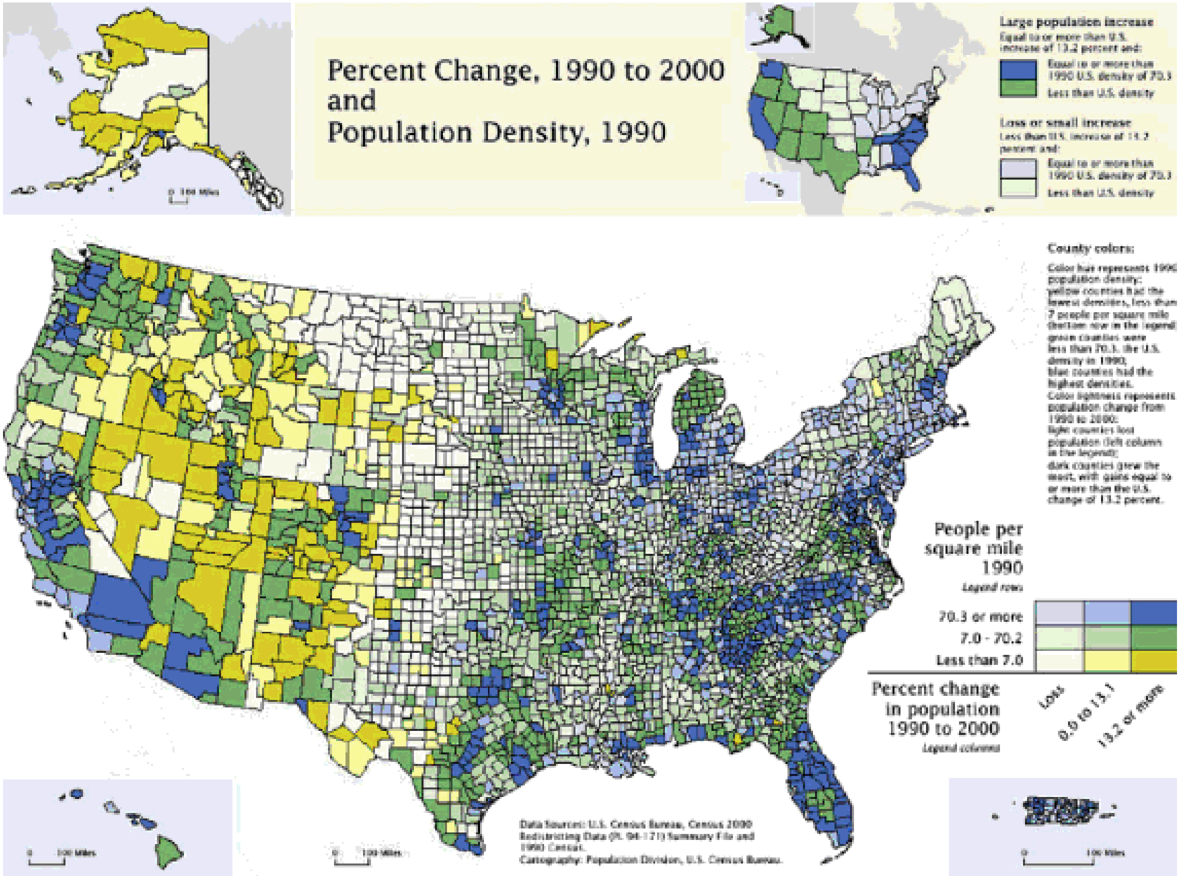


Figure 16 *Map representing bivariate coloring using saturation and lightness for one variable and yellow-green-blue variation for the other variable. (Ware 2013. p. 135).*

The problem in the color sequences is that the low-saturated colors are difficult to distinguish. Bivariate maps are known to be challenging to read. According to Ware (2013, p. 135), we humans do not seem to be able to read color dimensions in a way that is highly separable. A good way to show two variables on the same map is to use, for instance, texture, different heights, or another variable for one variable and color for the other. This way, the dimensions become more distinguished. Ware also mentions another alternative for separate dimensions by using computer graphics techniques such as artificial height and shading with artificial light source.

3.7.3 Multiform Bivariate Matrix

A Multiform Bivariate Matrix is a generalization of a scatterplot matrix consisting of univariate visualizations. This technique is not a complete scatterplot matrix. It applies another technique that suits bivariate data, which is a bivariate map. Histograms are situated along the diagonal axis. The histograms represent the number of values in each class of an attribute. On both sides of the diagonal axis the area of the matrix consists of bivariate visualizations. In this analysis, the upper side of the diagonal axis consists of scatterplots and the lower part consists of bivariate maps.

The scatterplot is based on the Cartesian coordinate system and is one of the earliest and most widely used technique in visualizing data and correlation between two variables. Items are represented in the scatterplot as circles or points. Multidimensional visualization tools allow giving graphical properties to data dimensions such as location, color, shape and size, which can vary (Figure 17). Scatterplot visualizations often provide mapping graphical properties for data items, which allow the user change the visual appearance dynamically. This, however, is not concretely related to the data dimensions or visual exploration. (Ward et al 2010, p. 35; Elmqvist et al. 2008, p. 1141.)

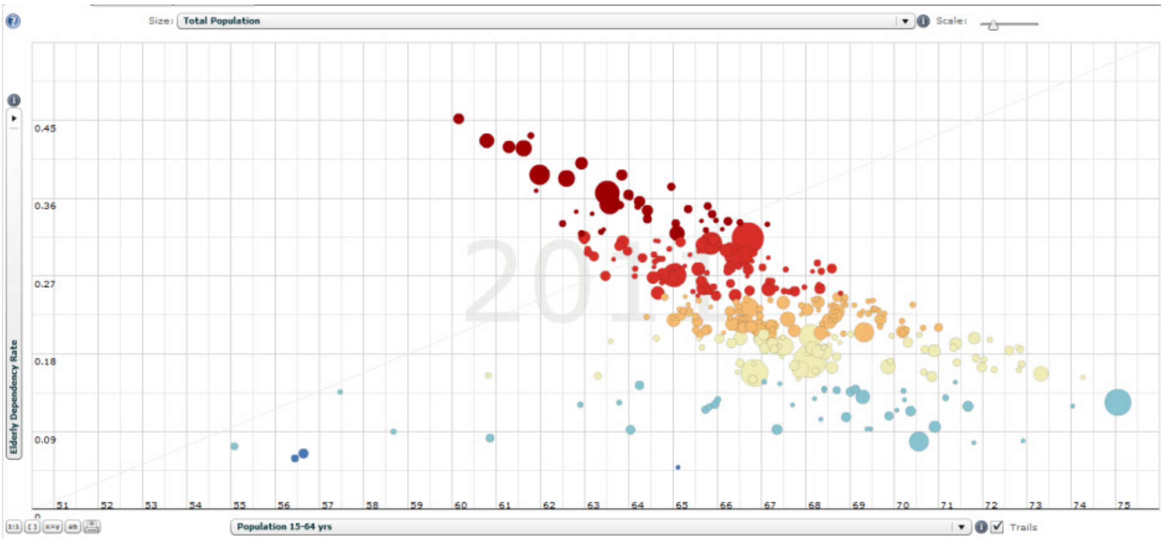


Figure 17 *Scatterplot presenting negative correlation between the population of 15–65 years and elderly dependency rate in 2011. (OECD 2014).*

Correlation between the variables can be positive (increasing from lower left corner to upper right corner), negative (decreasing from upper left corner to lower right corner), null (no correlated), or random (sparsely scattered). A trend line, also called the linear regression or best fit, can be drawn to depict the correlation between the two variables.

The usefulness of scatterplot matrices lies in their effectiveness in exploratory multivariate data analysis and they can be enhanced to provide more information (Carr et al. 1987, p. 424). Scatterplot matrix is the most common approach in the case of multiple displays. The matrix consists of grids of scatterplots, with the grid having N^2 cells, where N is the number of dimensions. Every pairwise plot is shown twice (Figure 18). They differ by a 90 degree rotation. The order of the dimensions or attributes is usually the same for the *x*- and *y*-axis. (Ward et al. 2010, p. 239.)

Scatterpolot matrix suffers from overplotting of points when plotting a large dataset into the matrix, in which each plot has a fixed dimension. The area of the screen is limited and, secondly, the closely neighbored points may fall in the same screen pixel. A massive overplotting of points may easily result in data loss, and make patterns and relationships difficult to identify. Overplotting at different pixels might include data that have some other patterns. Some pixels can have substantial overplotting points, while others have less. (Zhang et al. 2003.)

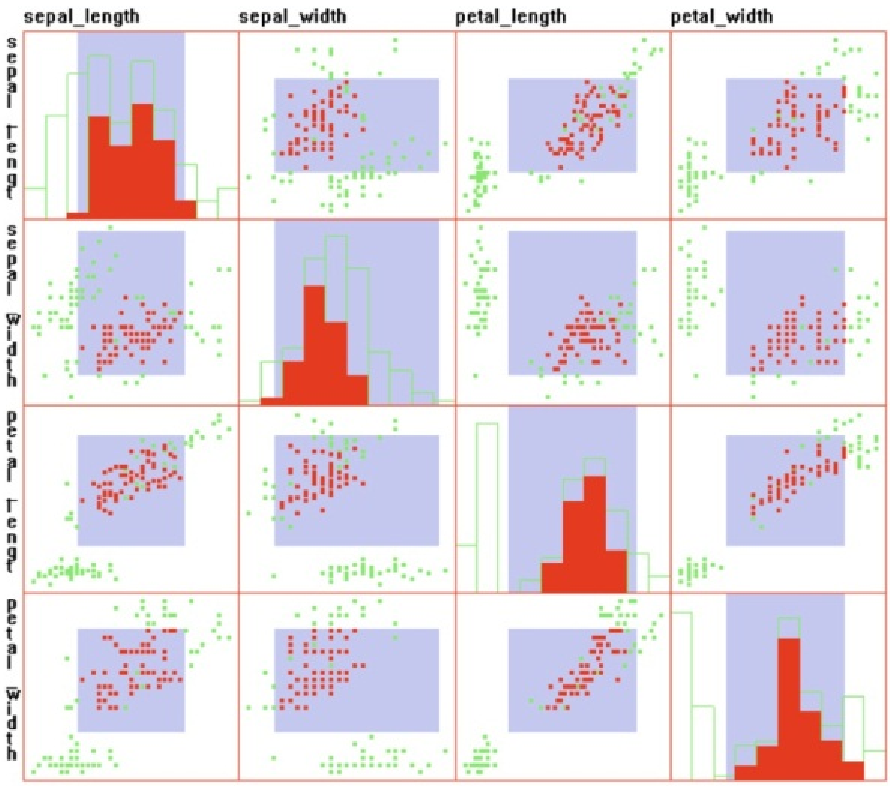


Figure 18 *Scatterplot matrix with histograms in the diagonal axis showing the highlighted data points. (Ward et al. 2010, p. 238).*

Another problem lies in the distributed datasets, which require a great amount of storage space. When the dataset is large, even transferring the subset through the network tends to become infeasible for some datasets. Plotting and computation times are still long and slow down the speed of response, which affects the user interaction. (Zhang et al. 2003.)

User interaction has an important role when exploring data with scatterplots. Large datasets provides the opportunity to explore densities and clutters, which can be detected in scatterplots. Techniques and tools working with scatterplot matrices need to move flexibly among density exploration, subset selection and comparison. One approach is data sampling which, according to Carr et al. (1987, p. 424), reduces display problems and provides cross-validation opportunities. However, outliers and other fine patterns, which are in the focus of interest, may be omitted by random sampling. Other solutions for overplotting problems are, for instance, 3D bivariate histograms, histogram with grayscale density or contour plot, and kernel density plots with perspective views. Brushing is a common tool for interacting with scatterplots, but still difficult in applying this for large datasets. (Zhang et al. 2003.)

This Multiform Bivariate Matrix used in this analysis is presenting bivariate maps in the grid cells on the lower side of the diagonal line. The upper grid cells of the diagonal axis are single scatterplots (Figure 19).

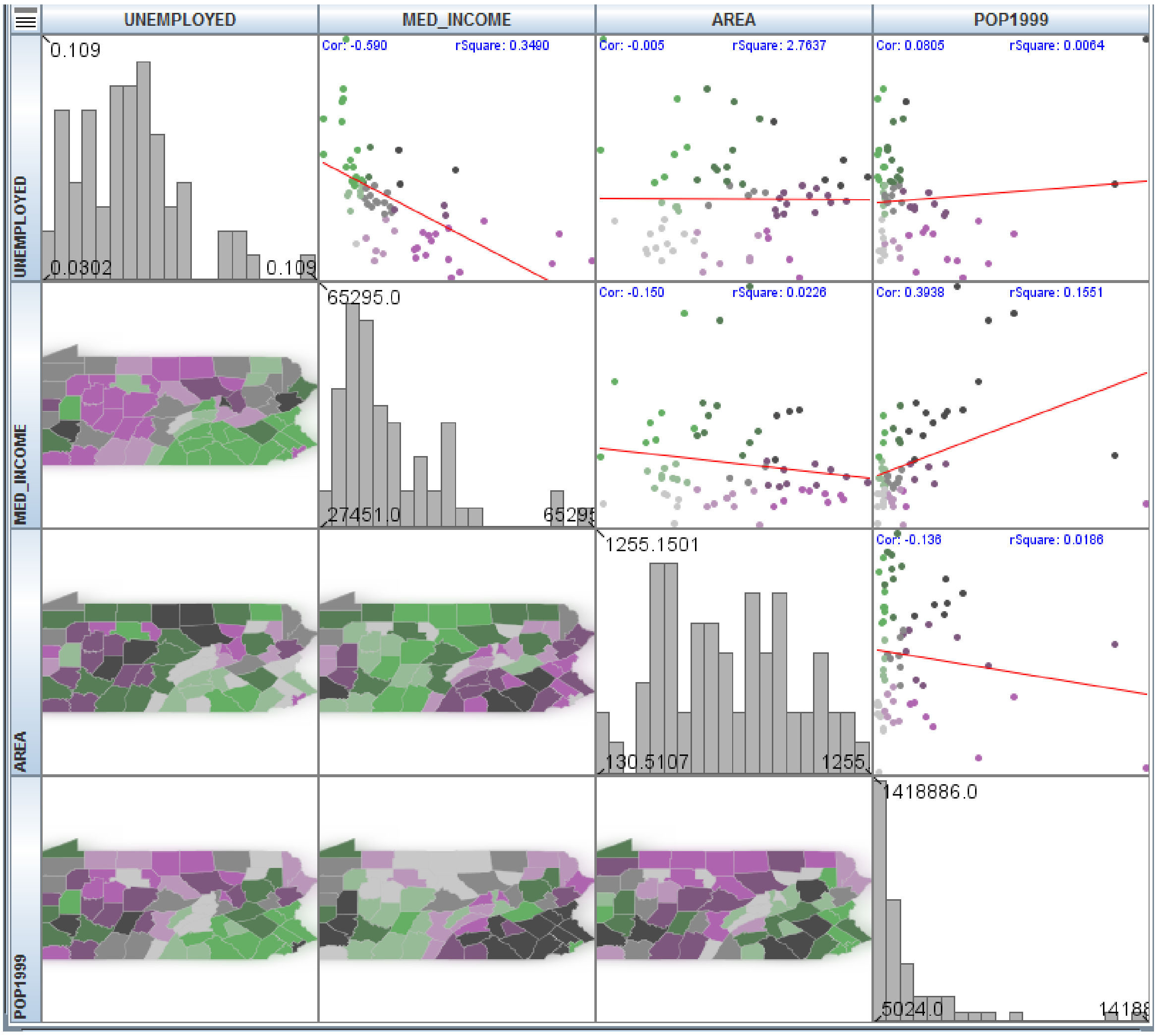


Figure 19 *An example of Multiform Bivariate Matrix with a bivariate map. Example data provided by GeoViz Toolkit.*

4 Material and data preprocessing

This chapter describes the data that was finalized for the analysis. The received data of the SAR organizations are briefly described with examples as well as the secondary weather and wave data. Huge datasets commonly include irrelevant information. For the analysis, the relevant data items need to be extracted. After the data construction and extraction, the items need to be classified including sub-classes. The classification is based on numbers meaning that for a certain type of data item a numeric value is given.

This chapter starts with the description of the SAR data of the SAR organizations in section 4.2, 4.3 and 4.4. Then follow the description and process of dealing with the secondary data on weather and wave parameters, sections 4.5, 4.56, and 4.7. Section 4.5 describes the received meteorological weather data. Wave data is reviewed in section 4.6 and lightning strike data is introduced in section 4.7. After that, the process of distance and cost surface calculations is introduced in section 4.8 and data preprocessing and classification are described in sections 4.9 and 4.10

4.1 Description of the attribute data

The primary data attributes needed in this study are presented in Table 1. The SAR organizations have some minor differences in keeping statistics on their missions. If the data is not available, it is not considered in this study. All the available attribute data for this analysis was gathered in an Excel 2010 table file. The information concerning the boating incidents was provided by the SAR organizations. In this analysis, the used SAR data columns are the following: incident type, type and size of the boat, name of SAR vessel, X and Y coordinates (WGS84) of the location, name of the municipality where the accident happened, date of the accident, time of the received alarm, accident quality and type of the accident.

Table 1 *The primary data attributes provided by the SAR organizations.*

	FBG (RVL)	FLI (SMPS)	FRS (PELA)
Incident type	x	x	x
Boat type	x	x	x
Boat class	x	x	x
Boat size	x	x	x
Location name		x	
Municipality		x	x
Coordinates x, y (WGS84)	x	x	x
Coordinates x, y (KKJ)			x
SAR stations	x	x	
Performing SAR station		x	
Date (dd.mm.yy)	x	x	x
Alarm time	x	x	x
SAR vessel		x	x
Accident quality	x	x	x
Accident factor	x	x	x

The accident quality categorizes the accident according to the accident factor. The accident quality includes factors that can be classified as *Unknown*, *Accident/Emergency*, *Technical* or *Other* (Appendix 3). The type of the accident is used for more detailed analysis. The accident type defines the factor of the accident. If some information (or part of it) was missing or clearly incorrect, e.g. the name of the municipality, it was checked and if possible, the actual information was given.

However, the available data could not fulfill all the required attribute information for every case. For instance, the name or ID of the SAR vessel was only available for FLI and FRS vessels. If the information was not available, it was left blank. All the data of every SAR organization was placed between the years 2007 – 2012. The date of the incident, also the day number, week number, weekends and national holidays were derived from the date information.

The data table includes information related to the incident, mission target (Boat type and size) and location (coordinates) (Figure 20), weather attributes (Figure 21), SAR organization, stations, and distanced (Figure 22), temporal facts, and type of the accident for further analysis (Figure 23).

	A	B	C	D	E	F	G
1	IncidentType	BoatType	BoatClass	BoatSize	Municipality	CoordLatY	CoordLonX
2	Rescue	Recreational vessel	motorboat	size unknown	Kotka	60,5006315	27,0325273
3	Rescue	Recreational vessel	motorboat	size unknown	Helsinki	60,1454536	24,9868636
4	Assistance	Recreational vessel	motorboat	size unknown	Helsinki	60,1700003	24,9665132
5	Rescue	Fishing vessel	other	< 15 m	Pyhtää	60,4193333	26,7541667
6	Rescue	Recreational vessel	motorboat	size unknown	Kirkkonummi	60,0249028	24,4136893
7	Rescue	Recreational vessel	motorboat	size unknown	Loviisa	60,3145837	26,5462573
8	Rescue	Recreational vessel	motorboat	size unknown	Porvoo	60,1862942	25,5645029
9	Rescue	Recreational vessel	motorboat	size unknown	Kotka	60,4399563	26,9235613
10	Rescue	Recreational vessel	motorboat	size unknown	Porvoo	60,1961845	25,794705

Figure 20 *An example of the data table showing the incident, vessel and location data provided by SAR organizations.*

H	I	J	K	L	M	N	O
WaveHeight_m	WaterTemp_C	AirTemp_C	WindSpeed_m/s	WindDirection_deg	Precipitation_mm	Cloudiness	LightningStrike
1.39	5.00	2.9	8.0	170.0			0
1.92	4.50	5.1	5.3	233.0	0.1	8	0
1.08	4.00	-8.0	2.3	37.0		1	0
0.51	3.75	-10.2	0.0	0.0			0
		5.7	5.5	255.0			0
		5.2	0.0	0.0			0
		-10.2	0.0	0.0			0
		3.0	6.0	240.0			0

Figure 21 *An example of the data related to weather, waves and lightning observations.*

P	Q	R	S	T	U	V	W
SAROrganization	SARStation	ClosestSMPSStation	DistClosestSMPS_km	ClosestRVLStation	DistClosestRVL_km	DistShore_m	DistOwnStation_km
RVL		Kotkan Meripelastusyhdistys r.y	7,652333	Kotka	7,500027	50,403756	
RVL		SMPS Meripelastusasema I	3,084452	Suomenlinna	0,401127	20,154606	
RVL		SMPS Meripelastusasema I	0,701838	Suomenlinna	3,18433	86,219815	
RVL		Kotkan Meripelastusyhdistys r.y	14,898369	Kotka	13,797122	191,656778	
RVL		Suomen Meripelastusseura Porl	4,034384	Porkkala	8,054572	64,458475	
RVL		Suomen Meripelastusseura Lovi	24,759947	Kotka	30,067827	46,376561	
RVL		SMPS Porvoon Meripelastajat - I	17,348923	Glosholmen	15,336985	495,074824	
RVL		Kotkan Meripelastusyhdistys r.y	3,697588	Kotka	2,596343	123,610394	
RVL		SMPS Porvoon Meripelastajat - I	22,588221	Glosholmen	2,715094	481,666822	
PELA		SMPS Porvoon Meripelastajat - I	10,105718	Glosholmen	24,596174	54,700257	

Figure 22 *An example of organization and distance-related data.*

X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
Date	DayNumber	WeekNumber	Month	Year	AlarmTime	Weekend	Holiday	SAR_Vessel	AccidentQuality	AccidentFactor
1.1.2007	1	1	1	2007	16:33	No	Yes		Accident/Emergency	grounding
16.1.2007	2	3	1	2007	8:54	No	No		Unknown	unknown
22.1.2007	1	4	1	2007	11:11	No	No		Technical	breakdown
25.1.2007	4	4	1	2007	7:51	No	No		Technical	breakdown
29.3.2007	4	13	3	2007	14:08	No	No		Unknown	unknown
31.3.2007	6	13	3	2007	8:36	Yes	No		Accident/Emergency	drifting vessel or object
1.4.2007	7	13	4	2007	16:34	Yes	No		Unknown	unknown
2.4.2007	1	14	4	2007	9:09	No	No		Accident/Emergency	MOB
6.4.2007	5	14	4	2007	20:20	No	Yes		Accident/Emergency	getting lost or uncertain spot

Figure 23 *An example of temporal data, SAR vessels (if given), and accident quality and accident factor.*

4.2 The SMPS data: SAR incidents and SAR stations

The FLI (in Finnish: Suomen Meripelastusseura – SMPS) compiles information and statistics of their missions in a system called MESSI (Venäläinen & Sonninen, 2013 p.3). The data is in an Excel 2010 table format with all the missions that took place both in seawaters and inland waters. In this analysis, the target groups were only the missions of twelve FLI stations located at the coast of the Gulf of Finland: Hanko, Tammisaari, Bågaskär (Inkoo), Inkoo, Porkkala (Kirkkonummi), Espoo, Helsinki (two stations), Porvoo, Loviisa, Kotka and Hamina (Table 2 and Figure 24).

Table 2 *SAR stations of the FLI provided with location and coordinates (WGS84).*

SAR Station	Location	Lat (X)	Lon (Y)
Hangon Meripelastajat	Hanko	59,8250	22,9450
Tammisaaren Meripelastajat	Tammisaari	59,9717	23,4317
Inkoon Meripelastajat	Inkoo	60,0400	24,0067
Bågaskärin toimintakeskus	Bågaskär, Inkoo	59,9312	24,0136
Porkkalan Meripelastajat	Kirkkonummi	60,0167	24,3483
Espoon Meripelastajat	Espoo	60,1539	24,7713
Meripelastusasema I	Helsinki	60,1667	24,9533
Helsingin Meripelastusyhdistys	Helsinki	60,1550	25,0850
Porvoon Meripelastajat	Porvoo	60,3367	25,5783
Loviisanseudun Meripelastajat	Loviisa	60,4417	26,2367
Kotkan Meripelastusyhdistys	Kotka	60,4583	26,9500
Haminan Seudun Meripelastusyhdistys	Hamina	60,5633	27,1800

Sations of the Finnish Lifeboat Institution located at the Gulf of Finland

Figure 24 *Map of the locations of the FLI SAR stations at the coast of the Gulf of Finland.*

First the data was checked for duplicates with the mission ID number (onnettomuusnumero). Then the coordinates were checked. Some of the coordinates had zero values. In some cases, the latitude and longitude coordinates had switched places. These flaws were corrected. Otherwise, all the cases without known coordinates were dismissed. All the coordinates for SMPS data are in the WGS84 (NGA 2013).

The data also includes the name of the SAR station that performed the mission. Based on this information the distance to the station could be calculated. In addition, the name of the SAR vessel, by which the mission was performed, was provided.

The FLI keeps record on the target groups of the missions consistently. The primary targets are recreational vessels, of which motorboats have the largest group (Figure 25). Motorboats are further categorized into three groups according to the size of the boat: less than 7 meters, 7–15 meters, and more than 15 meters. Besides motorboats, the boat types are classified to sailboats and other boat types. Other boat types include vessels of unknown type as well as non-traditional boats, such as jet-skies, kayaks, canoes, dinghies, sailboards, etc. This research is concentrated only on these three groups: motorboats, sailboats, and other.

For the more detailed exploration the size of the motorboats was taken into account. This is done only for the SMPS data provided by the FLI.

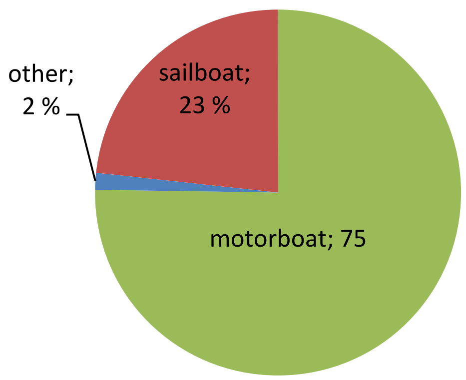


Figure 25 *Target groups of recreational vessels of FLI missions.*

4.3 The RVL data: SAR incidents and SAR stations

The FBG (in Finnish: Rajavartiolaitos – RVL) compiles their data to the database called RVT (rajavartiotoiminnan tietojärjestelmä). Also the data used in this analysis derives from this database. The coordinates are according to WGS84 coordinate reference system. No duplicates were found in FBG dataset based on the mission ID number (PETA tunnusnumero). Some coordinates with the value of zero were found and dismissed. Some of the locations were situated outside Finnish national border, mostly in Russia, but not many. It was not possible to relocate the locations, because the name of the municipality or actual place was not provided. It is assumed that the given coordinates were correct. Based on the given coordinates, the name of the municipality was checked and linked to the incident.

The FBG keeps track on the types of the boats. Motorboats, as in FLI cases, form the largest group (Figure 26). However, the motorboats are not categorized by their size as in the SMPS dataset.

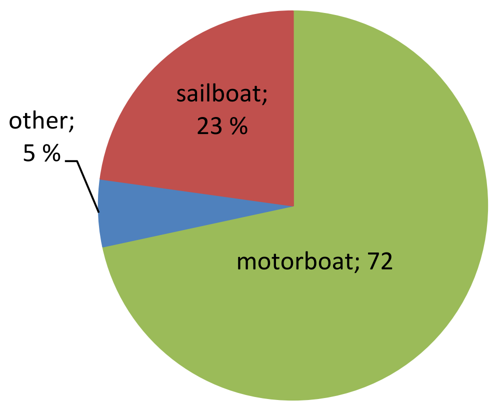


Figure 26 *Target groups of recreational vessels of FBG missions.*

The FBG has eight stations at the Gulf of Finland in Hanko, Tammisaari, Porkkala (Kirkkonummi), Suomenlinna (Helsinki), Glosholmen (Porvoo), in the city of Kotka, Haapasaari (Kotka), and Hurppu (Virolahti). Coordinates are provided in Table 3 and the stations are illustrated on a map in Figure 27.

Table 3 *SAR stations of the FBG provided with locations and coordinates (WGS84).*

Station	Location	Lat (Y)	Lon (X)
Hanko	Hanko	59,8108	22,9137
Tammisaari	Tammisaari	59,8902	23,5142
Porkkala	Kirkkonummi	59,9676	24,4099
Suomenlinna	Helsinki	60,1476	24,9818
Glosholmen	Porvoo	60,1908	25,8392
Kotka	Kotka	60,4546	26,9520
Haapasaari	Kotka	60,2860	27,1864
Hurppu	Virolahti	60,4903	27,7397

Stations of the Finnish Border Guard Located in the Gulf of Finland

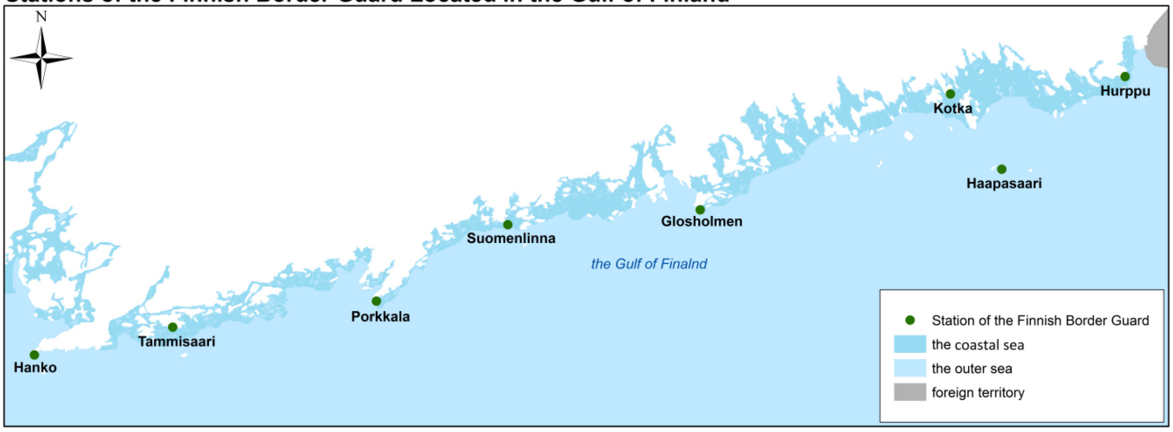


Figure 27 *Map showing the locations of the FBG SAR in the coast of the Gulf of Finland.*

4.4 The PELA data: SAR incidents

The FRS (in Finnish: *Pelastuslaitos* – PELA) keeps statistics on their resources and missions, which are recorded by regional rescue departments. The rescue data is restored in a system called PRONTO (pelastustoimen resurssi- ja onnettomuustilasto) by the Ministry of Interior (Venäläinen & Sonninen 2013, p. 3). The data was acquired from the Emergency Services College, which is in charge of the development and maintenance of the statistic data.

Quite many duplicates were found among the dataset, also based on the case ID number (onnettomuusselosteen numero). For an explanation, some records may have shown several times, for instance, if more than one SAR vessel had taken action in a mission.

The coordinates were provided mostly in the KKJ reference system (coordinates of the years 2007 – 2010 were all given in KKJ) (Finnish Geodetic Institute 2012). Those coordinates were transformed into WGS84 by using ArcGIS 10.1 application. Also the names of the municipalities and SAR vessels were provided in the data, but all information of the stations, such as location and the name, were missing. These data could not be completed in this analysis.

The boat types of the FRS accidents are unknown in this dataset. The FRS vessels are the size of equal or less than 10 meters, more than 10 meters or the size is unknown. Most of them are unknown.

4.5 Weather data

Besides the incident data, this research includes weather data, lightning strike observations and wave data that were provided by the Finnish Meteorological Institute (FMI). The weather data in this analysis derive from seven different weather observatories that are located along the coastal area of the Gulf of Finland. The observatories are located in Tvärminne (Hanko), Bågaskär (Inkoo), Mäkiluoto (Kirkkonummi), Kaisaniemi (Helsinki), Eestiluoto (Sipoo), Kirkonmaa (Kotka), and Koivuniemi (Virolahti). A map of the weather observatories is shown in Figure 28. Coordinates are provided in Table 4.

Table 4 *List of the weather observatories and their locations including coordinates (WGS84).*

Weather observatory	Location	Lat (Y)	Lon (X)
Tvärminne	Hanko	59,8440	23,2485
Bågaskär	Inkoo	59,9311	24,0141
Mäkiluoto	Kirkkonummi	59,9201	24,3493
Kaisaniemi	Helsinki	60,1752	24,9446
Eestiluoto	Sipoo	60,1246	25,2228
Kirkonmaa	Kotka	60,3852	27,0486
Koivuniemi	Virolahti	60,5273	27,6727

Figure 28 *Map showing the locations of the weather observatories at the Gulf of Finland.*

However, it should be noted that the observatories are not evenly located along the coast of the Gulf of Finland. For instance, there is a long distance between Eestiluoto and Kirkonmaa observatories. No weather observatory exists between Sipoo and Kotka municipalities. Therefore, the available data was extended to cover the neighboring municipalities as properly as possible; otherwise there would have been too many gaps in the final compiled dataset. This was done, in order to get the data as even as possible. But for this reason, the used weather data is not representing the exact conditions during the time of an incident.

Since the incidents are located in the coastal area of the Gulf of Finland, some very near the shore, some in bays and others further away in the open waters, the surrounding conditions can be very local and thus inconsistent and varying in the neighboring areas. For this part, expert knowledge could have been applied to analyze the possible weather phenomena that are typical for the coastal areas. Alternatively, atmospheric hindcast models could have been applied to determine the weather-related conditions more accurately. These options were not used in this part of the analysis.

Depending on the weather station, the observations are taken every 10 minutes, every hour or every three hours. The weather data included information on the air temperature (ºC), average wind speed (m/s), wind direction (deg), total cloud cover (0…9) and precipitation (mm). Also the station ID, date and sampling time were provided. Only precipitation was delivered in a different file from the other weather data and is taken every hour. However, not all the information is available for every observatory. Therefore, the data of a neighboring municipality is used whenever needed.

The observations of air temperature, wind speed and wind direction were quite well covered. The air temperature was given in Celsius degrees and the average wind speed in meters per second (Figure 29). The wind directions was given in degrees, where 360º represents the direction up North, 90º to the East, 180º to the South, and 270º to the West. Zero degrees indicates no wind at all. Cloudiness was given in part of eighths. Zero represents clear sky. Value 8 indicates that the sky is completely covered by clouds. Value 9 means that the observation could not be measured due to fog or for similar reasons that prohibited the measurement.

	A	B	C	D	E	F	G
1	Station_ID	Date	Time	Average wind speed	Wind direction	Temperature	Cloudiness
2	304	20070101	0	3.5	286.0	2.3	1
3	304	20070101	10	5.0	286.0	2.3	3
4	304	20070101	20	4.3	287.0	2.2	5
5	304	20070101	30	5.0	287.0	2.2	7
6	304	20070101	40	5.1	292.0	2.2	8
7	304	20070101	50	5.4	296.0	2.2	8
8	304	20070101	100	4.4	297.0	2.1	8
9	304	20070101	110	5.1	294.0	2.1	8
10	304	20070101	120	4.5	298.0	2.1	7

Figure 29 *(edited) An example of data table of weather data. This data derives from Kaisaniemi observatory in Helsinki.*

Information of precipitation and cloudiness was missing from many observatories. Precipitation was given in millimeters. The value -1.0 means that there has been no rain at all. Zero value indicates that there have been some observations of rain but significantly less than the accuracy of the observed measurement. Data cells for those areas that could not be filled up with any available values were left blank. Table 5 lists the available data from each weather observatory including the quality of the data according to the personal judgment (1 – 5; 1 = poor, 5 = very good).

Table 5 *Data parameters provided by the weather observatories of the FMI.*

	Tvärminne	Bågaskär	Mäkiluoto	Kaisaniemi	Estiluoto	Kirkonmaa	Koivuniemi
Sampling time	10 minutes	10 minutes	10 minutes	10 minutes	10 minutes	3 hours	10 minutes
Wind speed (m/s)		x	x	x	x	x	x
Wind direction (deg)		x	x	x	x	x	x
Air temperature (°C)	x	x	x	x		x	x
Total cloud cover (0–9)		x		x			
Precipitation (mm) (1-h average)	x		x	x			x
Quality (1–5)	1	4	3	4	2	4	3

4.6 Wave data

The FMI has four buoys for measuring waves in the Baltic Sea. They are located in the Northern Baltic Proper, off Helsinki, in the Bothnian Sea and in the Bay of Bothnia (Figure 30). In this analysis, only the data that derives from the buoy of Helsinki is used.

The wave data of the FMI includes the following parameters: the significant wave height (m), the peak period (Hz) (which is the period with the highest energy level), the direction of arrival at peak period (deg), the direction of deviation (deg), and water surface temperature (ºC). In this analysis, only the significant wave height and water surface temperature were used. The data are taken every half hour during years 2007−2012. Some gaps between the dates were found in the dataset. Those periods which could not be provided with wave data were left blank.

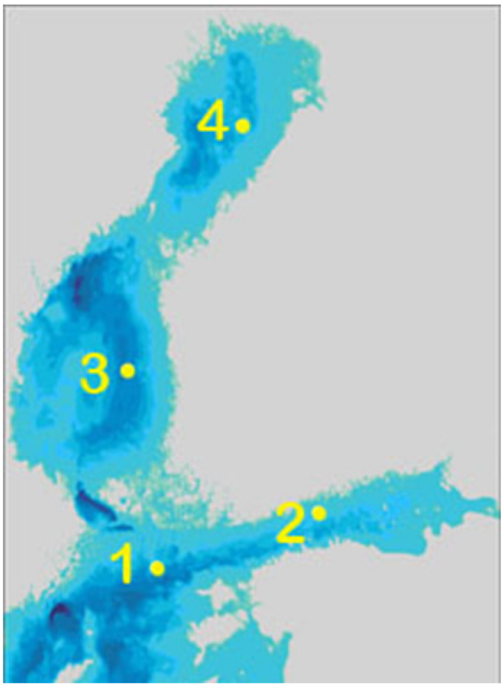


Figure 30 *Buoys of FMI in the Baltic Sea: 1. the Northern Baltic Sea, 2. Helsinki, 3. the Bothnian Sea, 4. the Bothnian Bay. (Ilmatieteenlaitos 2013). Source: http://ilmatieteenlaitos.fi/image/image_gallery?uuid=2339fad8-988d-4dd7-b417-c3d61a70b33d&groupId=30106&t=1339752933865 [Referred 28.10.2013].*

The buoy off Helsinki, however, is located that far from the main land that it cannot expected that the measured significant wave height and surface temperature is representing the actual water characteristics around the incidents, because most of the cases are located near the coast of the main land. According to Leppänen et al. (2012, p. 16), coastal waters get warmed-up faster than the waters in open sea areas because of their low level of depth (Figure 31). The sea temperature gets higher from the surface due to the solar radiation. Also wind action, ocean currents, evaporation, precipitation, and river inflows have an impact on the variation of the temperature of seawater surface (Mitina 2006, p. 285). Therefore, the water temperature of the surface level can be several degrees higher than in the lower levels of depth during summer time.

The annual variation range of the temperature of open seawater surface varies approximately from a bit under 0 ºC (in winter) up to +19 ºC (in summer) (Leppänen et al. 2012, p. 32). The water surface temperature, measured by the FMI, was taken 0,35−0,45 cm below the water surface. The water level at the coast varies approximately between -130 cm and +200 cm. (Ilmatieteenlaitos 2013.)

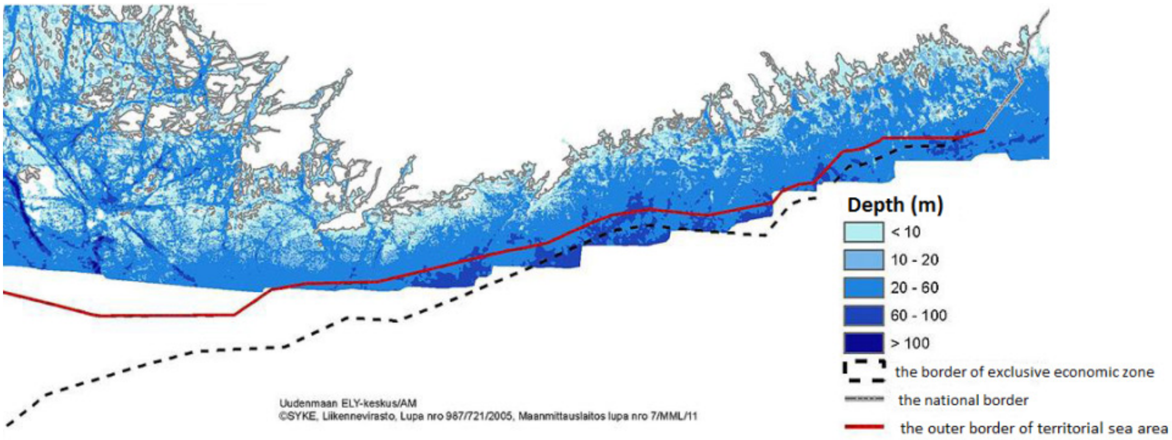


Figure 31*(edited) A rough depth classification of the Gulf of Finland. (Map: UUDELY. MML, the Finnish Land Survey 2006). (Leppänen et al. 2012, p. 17).*

The measured wave data of Helsinki buoy was given to all incidents in this analysis. Since the data is not representing the coastal water characteristics, it can be assumed that the wave height near the coast cannot be higher than at the Helsinki buoy. The water surface temperature, correspondingly, can be expected to be higher near the coastal areas based on the facts described above. The assumed water temperature cannot be less than the temperature at the buoy of Helsinki. However, these assumptions should also be analyzed based on expert knowledge or more detailed wave models.

4.7 Lightning observations

The lightning detection method of the FMI is radio positioning. The key parameters are time and location data. The time is registered as UTC time and the coordinates are in WGS84 format. (Ilmatieteenlaitos 2013.) This research focuses only on the occurrence of possible lightning observations during the same time as an incident took place within a certain area. As the time data for lightning observations is according to the UTC time, the time was transformed to match the local Finnish time, also taking both the summer (UTC + 3) and the winter (UTC +2) time into account.

Over the entire observed area (the Gulf of Finland), a so called Fishnet layer was placed on top with ArcMap 10.1. The net divides the observed area into 10 km x 10 km quadrangle grids cells. The lightning strikes and incidents were investigated within the grid cells.

Another method for checking the lightning strikes, besides Fishnet, would have been to create buffers around each incident. Fishnet method was used, because buffers, in this case, tend to overlap in densely clustered areas, which makes it hard to join and extract the 1ightning data with the incident. Number 1 indicates that there has been a detected lightning strike approximately at the same time during the day when the incident took place. Zero correspondingly means no detected lightning strikes within the area and its neighboring areas of an incident.

Lightning strikes are related to thunderstorms. The season for thunderstorms is short, from May to September. During that period most of the boating incidents tend to occur, since this period is ideal for boating. Mäkilä (2009, p. 15) mentions in her study that thunderstorm is the main cause of boating accidents after grounding. According to Mäkilä (2009, p. 15), a thunderstorm might occur instantly and be unnoticed. The warmer the air temperature is the greater are the preconditions for the emergence of a thunderstorm. Addition to this, only adequate humidity and a vertical flow (an unstable troposphere) are required to cause a thunderstorm. (Ilmatieteenlaitos 2013.)

This analysis is not focused on whether a lightning strike was the main cause of the accident. The interest is in whether the surrounding conditions refer to thunderstorm. Based on the reconditions of a thunderstorm, an adequate precipitation, wind speed, air temperature and cloudiness will be needed, and those parameters were considered.

4.8 Distance calculations

The distances used in the table columns are all calculated by using the *Cost Path* method provided by ArcGIS 10.1 under *Spatial Analyst Tools*. The *Cost Path* tool determines the least-cost path from the destination source to the source location (Esri 1995–2012) (Figure 32). For the *Cost Path* method, two other raster layers, *Cost Distance* (Figure 33) and *Cost Back Link* (Figure 34), had to be created first.

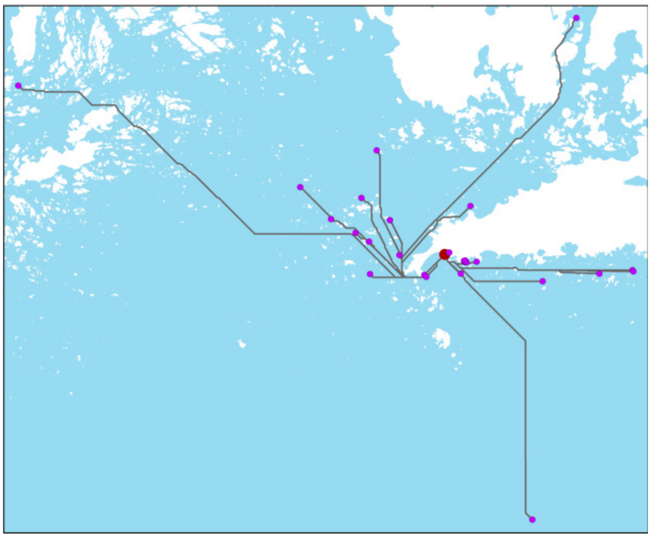


Figure 32 *Cost Path method used in calculating distances to the FLI Hanko station.*

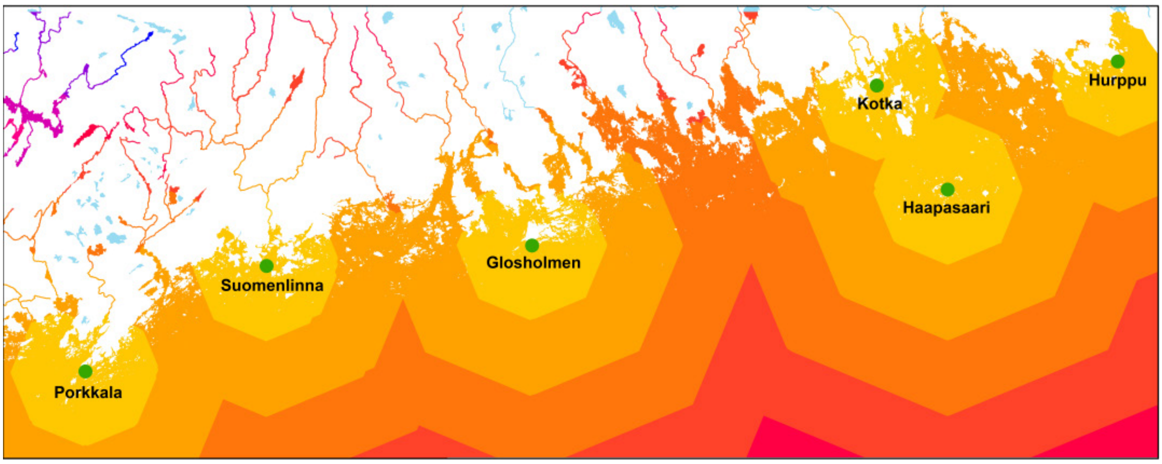


Figure 33 *An example of Cost Distance calculation for stations of the FBG.*

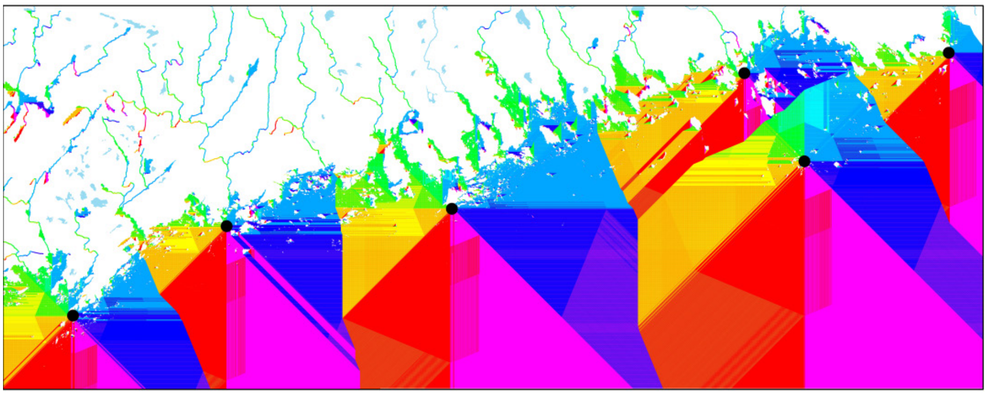


Figure 34 *Back Link raster for the stations of the FBG.*

The *Cost Distance* tool calculates the least accumulative cost distance for each raster cell over the cost surface to the nearest source. The cell value is determined by the cell size multiplied with the cost value. The back link raster is used for retracing the least costly path from the destination to the source location over the cost surface. The back-link raster includes values of 0−8, which identify the direction (the next neighboring cell) along the least accumulative cost path from a cell to reach the source with the least cost. The directions are depicted in figure 35, where value 0 is the source cell (gray-colored). In Figure 34, the directions are depicted with colors. The source cells, in this case, are the locations of SAR stations. Cells that have *NoData* as a value are excluded from the set. (Esri 1995−2011; Esri 1995–2012.)

Figure 35 *Backlink positions. (Esri 1995−2011).*

The cost distance surface requires a source location, which in this case are the locations of the stations, and a cost raster layer as inputs. The cost raster layer represents the water area and identifies the cost of the travel path through each cell. The raster cell size is 10 m x 10 m. The points of the incidents and SAR stations are located over the cost surface. If some of them happened to be located on a land area, they were relocated on the water surface nearest to the original position.

The water-cost-surface was classified with the value of 1. *NoData* value was given the land area, so that the land area will be excluded. If the *Cost Path* tool was not able to measure the path, e.g. the incident was located too far away from the station, difficully situated or overlapping with a path or another incident, the route was measured manually.

The distance to shore was calculated by using the Join tool in ArcMap 10.1. The option "*Join data from another layer based on spatial location*" was selected. It joined the incident-point data layer to the land layer calculating the closest distance to the land.

There was no information of the station locations of the FRS in the dataset. No distances were calculated for missions that were performed by the FRS. Also the station information of the FRS was missing. Only the locations performed missions were provided. The RVL data was provided with stations (coordinates and name). However, it did not provide information on which missions the stations have been taking action. For this reason, the home station of an incident is unknown. The calculated distances are thus the distance to the nearest the FBG and the FLI station, and the distance to shore.

The SMPS data provides the most comprehensive data for distance calculation and for the dataset itself. All the incidents are recorded with the SAR stations that performed the mission. For the FLI cases also the distance to the performing station was possible to calculate.

The distance calculation process with ArcMap 10.1 is, however, time-consuming due to creation of several cost layers, as well as editing of locations of point data, manual check-up of path distances, and review of coordinate system. The coordinates in this process have to be transferred into the same system, KKJ in this case. Otherwise, the cost distance calculation process would have failed.

The distances to the stations are presented in kilometers. The distances to the nearest shore line are in meters. This selection was based without any specific criteria.

4.9 Data cleaning and integration

The original data was compiled and maintained in Finnish. To ease the analyzing process, the Finnish nominal data is translated into corresponding English terms. This way it was easier to deal with the data. Before this step, the data had to be checked and compared to find out between every organization how the data was recorded.

For this analysis, additional weather and wave data was provided by the FMI. The data files were big by their size, approximately 300 000 rows in Microsoft Excel (~55 MB). The data derived from seven different weather observatories. For the data extraction, the format of the date and time data had to be modified to match the format of the final data table.

Once the data was processed for the data extraction, VLOOKUP function was used with Microsoft Excel 2010. The function searched for the values in the first column of a table array and returned the values in the same row from another column, even from a different file. For this extraction part, if needed, the time durations of incidents were rounded up to the nearest ten minutes, in the case of weather data extraction, because the weather observations were mostly taken every 10 minutes.

This process was also done to extract the wave data, but to the nearest half hour. In the case of precipitation data extraction, the time is rounded to the nearest hour. For lightning strikes this method was not used.

Despite the rounding of time values, some gaps remained in the final dataset. The main reason is that there was no data available for that particular period of time or the provided data did not cover the particular area where data was needed.

4.10 Data classification and transformation

Techniques in GeoViz Toolkit deal with numeric values. The nominal values (names, places) were transformed to numeric values with Microsoft Excel 2010. The classification of variables and numeric values are listed in a table in Appendix 2.

Every station has, for instance, its own standard for the ID numbers and categories. For example, every case is provided with an ID number, as well as the type of the mission, accident quality and the cause of the accident. The applied ID numbers, however, are mostly not used because they are different among the stations.

This section describes the incident types, accident factors and other attributes.

4.10.1 Accident factors

The more detailed analysis is focused on missions of the FLI. Therefore, in the categorization of the factors of the accidents the numeric numbers are mostly given according to the ID codes used by the FLI (see Appendix 1). In this research the numbers for different factors are listed in a table in Appendix 2.

The FLI has ambulance and emergency factors listed as a separate category, but this analysis studies the factors in a more general level. All factors that are categorized as *Accident/Emergency* start with number 1. Correspondingly, the first code number of a *Technical* factor starts with 2, other factor starts with 3 and *Unknown* factor starts with 0. The codes are presented in a table in Appendix 3. These factors will only be included in the more detailed analysis of the FLI data.

Figure 36 shows a diagram that presents the percentage share of each factor in the dataset. In the chart, it is shown that the dominant group of factors is breakdown which is a technical factor. After the technical factors, the unknown factors form the largest group. One of the reasons is that the information is missing from the FRS data. The third most common factor is grounding. The chart in Figure 36 gives an overview of the factors in the whole dataset. The individual accident factors that are not common such as icing conditions will not be explored profoundly.

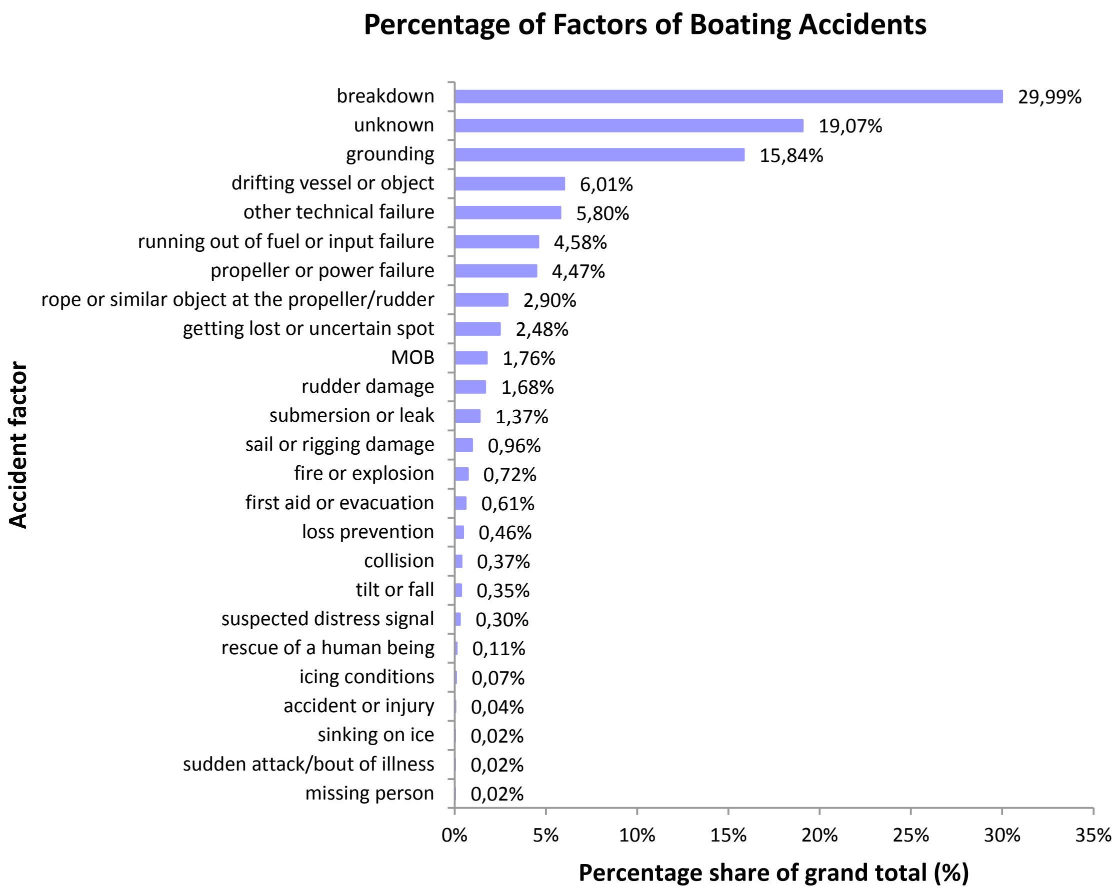


Figure 36 *The individual accident factors of the dataset in percentage.*

4.10.2 Search and rescue missions

At the beginning of the research, missions are visually mapped and the data statistics are visually illustrated by diagrams in Microsoft Excel 2010. Figures 37 and 38 are illustrations of all the cases classified by the type of the incident. Both charts (Figure 37 and Figure 38) show that *Rescue* class forms the largest incident group. For instance, the most missions performed by the FBG are recorded as a sea rescue mission (figure 39). The FLI typically performs missions for a need of assistance (figure 40). The FRS seems to be taking action mostly on missions in a case of an accident or emergency, and on those minor tasks that are common on land (Figure 41).

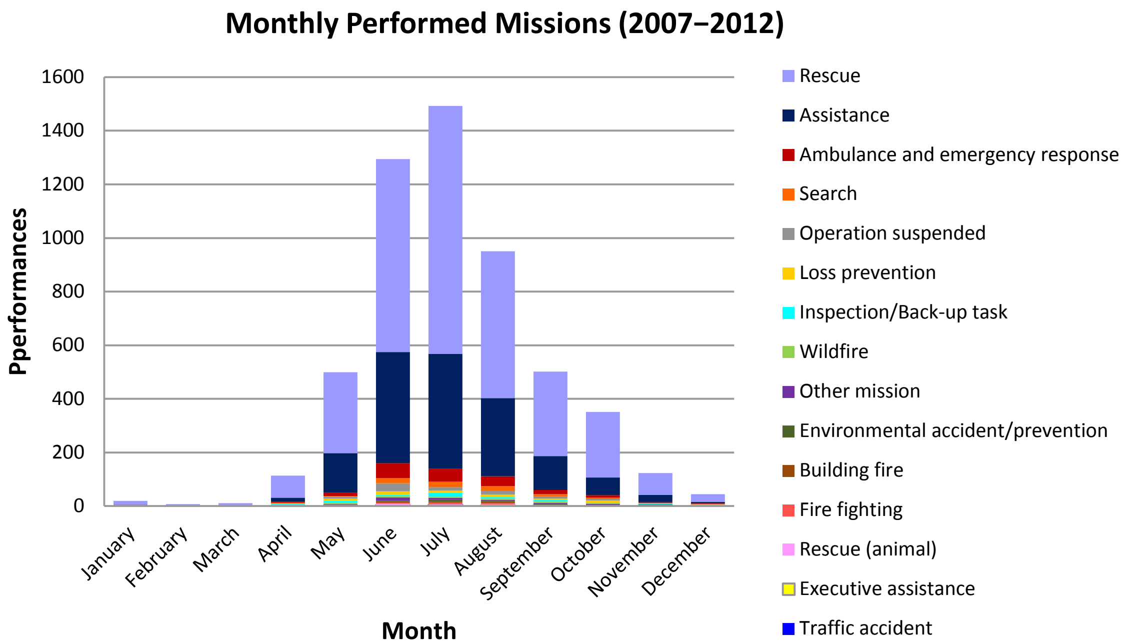


Figure 37 *Mission types and their occurrences during each month 2007–2012.*

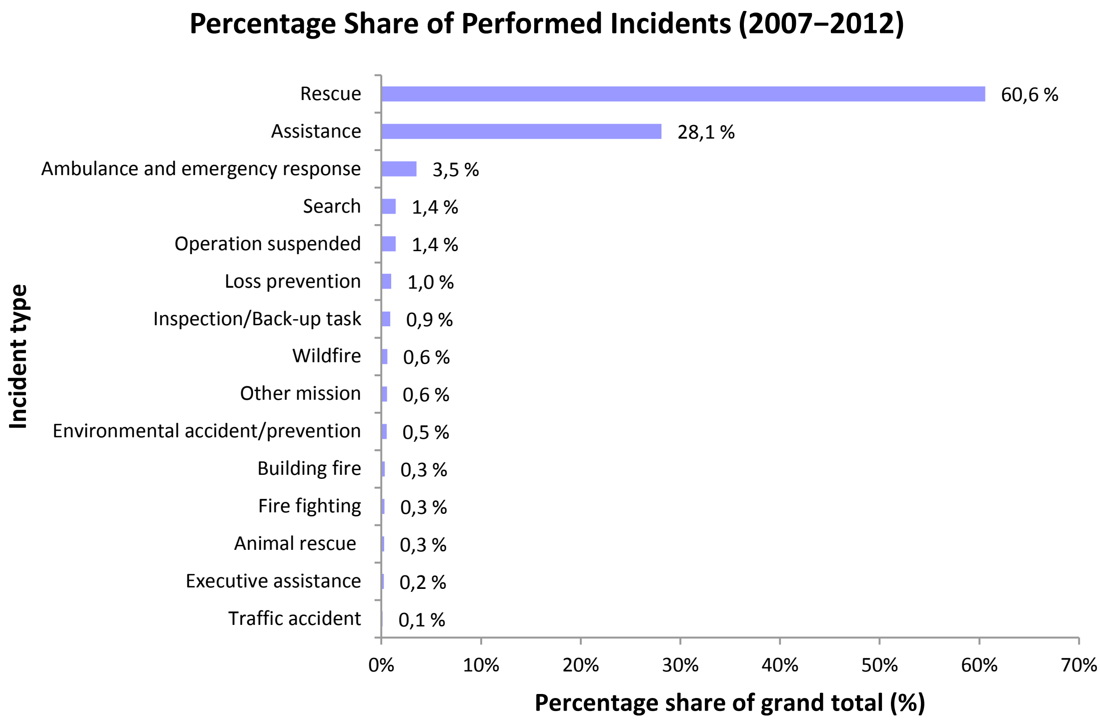


Figure 38 *Types of missions in percentage of the whole dataset.*

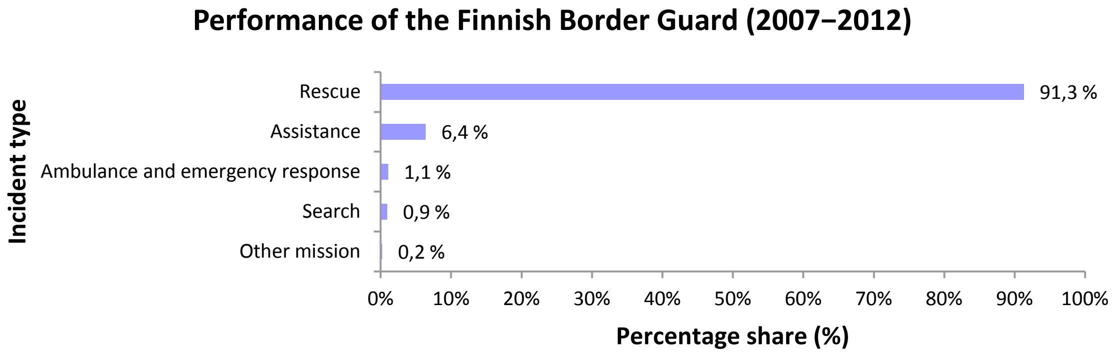


Figure 39 *Mission types of the FGB.*

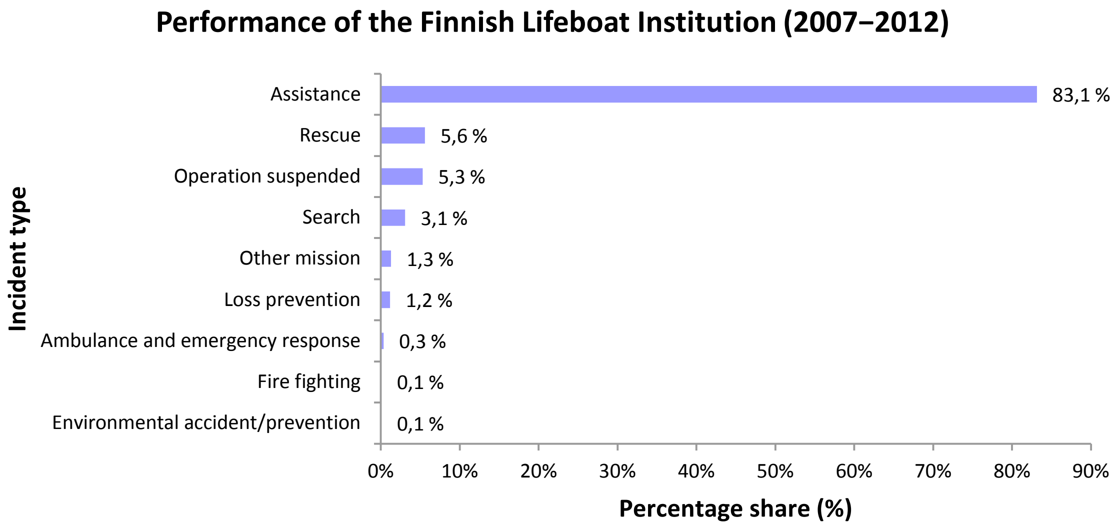


Figure 40 *Mission types of the FLI.*

Figure 41 *Missions types of the FRS.*

The problem in classifying the mission types is in the large number of classes (15 different types). The number of classes is reduced. The minor classes can thus be explored either separately from the major groups or as a one group consisting of the minor classes. This approach was taken into account, when dealing with numeric values (see Appendix 2).

Inspecting figures 37 and 38, the five major groups are 1 *Rescue*, 2 *Assistance*, 3 *Ambulance and emergency response*, 4 *Search*, and 5 *Operation suspended*. These groups are numbered from 1−5 starting from the largest group. To the group number 6 are included the minor classes that are 6.1 *Loss prevention*, 6.2 *Inspection/Back-up task*, 6.31 *Wildfire*, 6.32 *Building fire*, 6.33 Firefighting, 6.4 Other mission, 6.5 *Environmental accident/prevention*, 6.6 *Animal rescue*, 6.7 *Executive assistance*, and 6.8 *Traffic accident*. The order of the numbers is according to the amount of the performances, from the largest to the smallest. Note that *Wildfire*, *Building fire* and *Firefighting* make an exception. This is because these types are related to fire and thus classified so that they have similar code values. For example, value 6.3 can be indicated as a mission performing fire-related accidents.

The chart in Figure 42 shows the percentage share of each minor group for each month during the years 2007–2012. This diagram indicates the share of each month. One month represents the whole share. The ten minor groups appear during summer months from June to September. Inspection and back-up tasks (dark turquoise) occur throughout the year. The appearance of the rest of the missions is centered to the summer months.

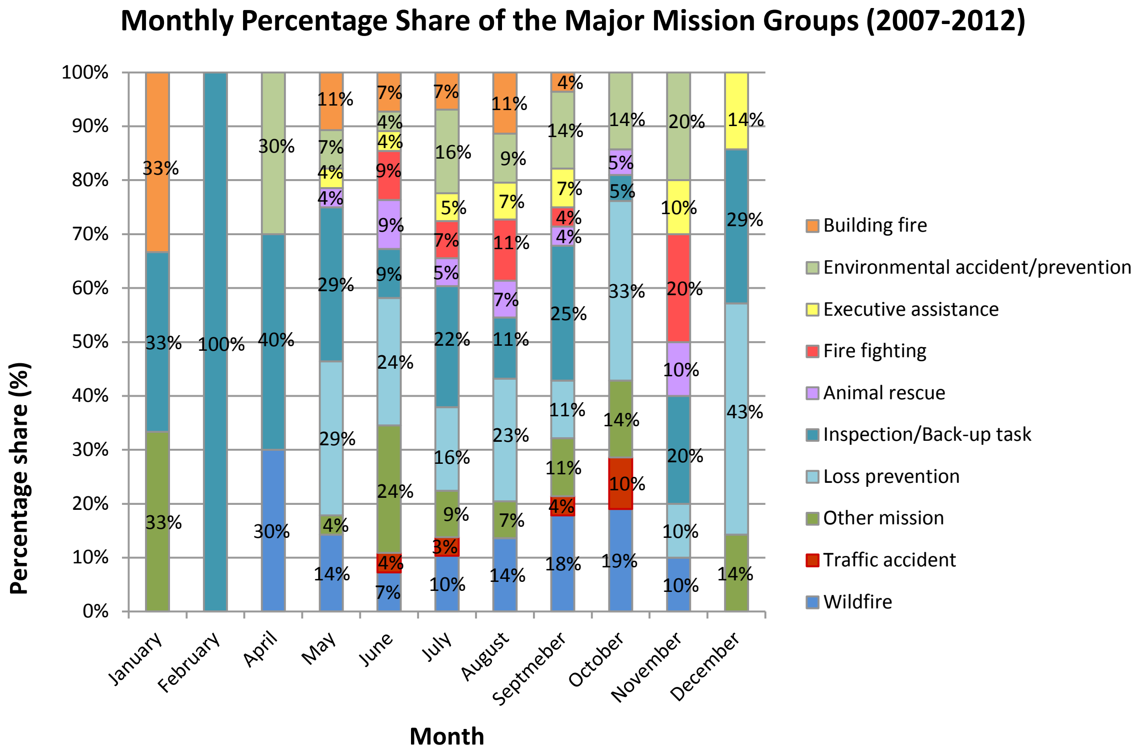


Figure 42 *Monthly shares of the minor missions in 2007–2012.*

Just like all the missions, including the dominant groups, the peak period of all mission types still seems to be June and July while inspecting the Figure 37 that was introduced earlier. The pattern of the distribution seems to be somewhat the same, despite the type of the mission. Figure 43 shows the percentage share of the five largest incident groups. *Rescue* missions are performed throughout the year as well as missions of *Assistance* and *Ambulance and emergency response*.

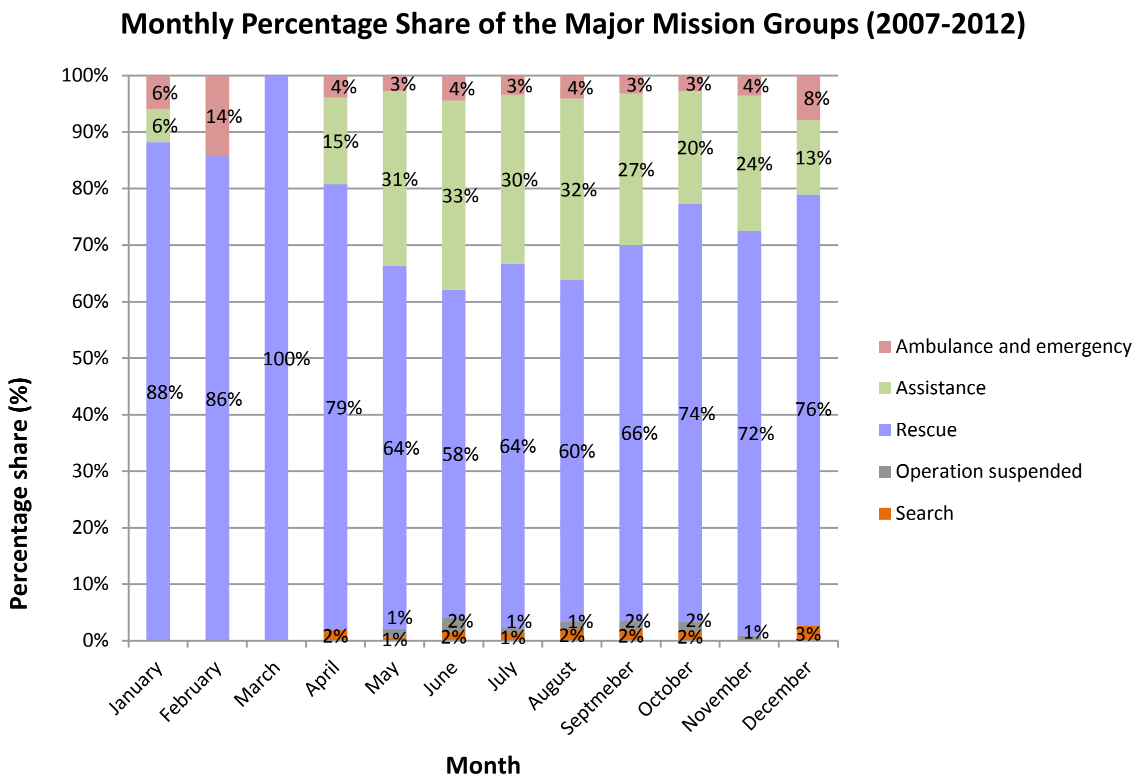


Figure 43 *Monthly shares of the major missions in 2007−2012.*

4.10.3 Classification of the rest of the attributes

The SAR stations of the FLI start all with number 2. The reason for this is that 2 indicates the organization itself (1 = The Finnish Rescue Services, 2 = The Finnish Lifeboat Institution, and 3 = The Finnish Border Guard). The same was done to the stations of the FBG which start with number 3.

The order of the station numbers start from the West to the East. The FLI has twelve stations at the Gulf of Finland starting from Hanko (1) to Hamina (12). The FBG has eight stations starting from Hanko (1). Hurppu (8), located in Virolahti, is the last SAR station of the FBG. The codes of the stations are listed in Appendix 2.

The same premise was applied to the municipalities. The Gulf of Finland coast consists of 15 municipalities. The numbers starts from Hanko (1) to Virolahti (15). Note that Pernaja (10.1), Loviisa (10.2) and Ruotsinpyhtää (10.3) are exceptions concerning the numbering. The reason is that Pernaja and Ruotsinpyhtää have been amalgamated with Loviisa since 2010. Missions performed in these areas, before the amalgamation came into effect, are recorded in the dataset according to the IDs that were being operative that time.

Those areas that are not part of any municipality located in the Gulf of Finland are named as *External* and valued with the number of zero (see Appendix 2). Figure 44 presents a chart, in which municipalities the cases appear. The majority are located in the metropolitan area, Helsinki and Espoo. This can be explained in many ways. For instance, the number of the population density is known to be the highest in that area. Consequently, also the number of local boaters probably is higher than in other area. Secondly, boaters can reach long distances with their vessels to other places if there is a national holiday or an annual event. Relevant is the fact that the metropolitan area is an attractive destination for boaters traveling long distances.

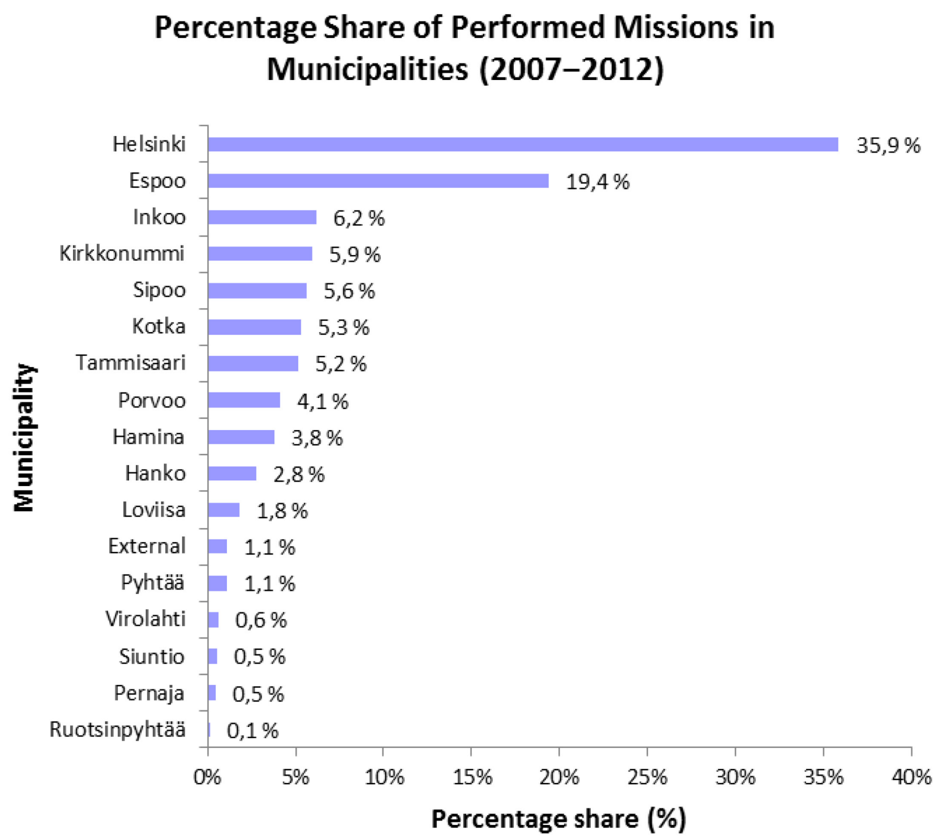


Figure 44 *Percentage of the missions in municipalities.*

In the data table is listed also the information whether the incident happened during weekend or national holiday, when the majority of people are free from work. This information is either “Yes” or “No” and assigned as numbers 1 (Yes) or 0 (No).

The FLI keeps track on the SAR vessels that take action on a mission. The SAR vessel names are listed in Appendix 4 (1/2). The list includes the home harbor and the municipality where the vessel has performed the majority of its missions. The vessels are numbered according to their home municipality. If the home municipality is unknown or the home municipality does not locate in the Gulf of Finland, the given code number of the municipality is applied. However, some exceptions are included, which must be taken into account in the analysis. It is recommended to check the vessel from the table list.

The bar chart, attached in Appendix 4 (2/2), shows the statistics of the missions performed by the SAR vessels of the FLI.

4.11 Mapping

The background maps for geographic reference of the images are acquired form PaITuli geospatial data service. The downloaded data derive from the National Land Survey of Finland (MML Maanmittauslaitos) and Finnish Environmental Administration (SYKE Suomen ympäristökeskus). This data and data elements are edited and illustrated by using ArcMap 10.1.

5 Results

The main task of this research was to investigate the SAR missions with additional weather and wave data as well as compare the SAR organizations. The aim was to find answers to the following questions:

- What are the general conditions when a certain type of an incident tends to happen?
- Under what conditions do the SAR organizations perform their missions? Are there any difference?
- Which visual analysis technique is suitable in discovering the patterns? Can we find any dependencies or outliers?

The applied techniques are common visual analytic tools provided by GeoVis Toolkit: parallel coordinate plot (PCP), GeoMap (or bivariate map), and Multiform Bivariate Matrix. The functionality of GeoViz Tookit and the techniques is not reviewed in this chapter. The characteristics of the toolkit environment are more thoroughly described in Appendix 6.

The result images are snapshots of the results. The selected images represent the main examples and findings of the explored data. The figures are selected in accordance with the hypotheses that support the exploration. The hypotheses are such as:

- "Wind speed and wave height are high during a bad weather",
- "Most missions are performed during the day hours. What is the pattern during the night and late-evening hours ?", and
- "The radius of action is varying between the SAR organizations, meaning there are differences in travel distance between the missions and the SAR station and between the missions and the shore."

These hypotheses are applied to all three SAR organizations together. More detailed analysis is applied to the FLI data, which includes the data of accident factors as well as boat types and boat sizes and information about SAR vessels.

The answers to the first two research questions are based on the result images of section 5.1. Sub-section 5.1.1 studies the SAR organizations and their activity in different the wave and wind conditions and the appearing types of the incidents. Sub-section 5.1.2 refers to the second hypotheses, where the appearance of the missions and proportions of different incident types in different areas are compared. The travel distances between the incidents and stations and distances to the shore are compared for each SAR organization in sub-section 5.1.3. This section refers to the third hypotheses, where the radius of action of each SAR organization is investigated. Section 5.2 is the part of detailed analysis on the FLI data. Section 5.3 presents some examples where data of SAR vessels are applied.

The next sections are referring to the first and second research questions. The focus of the study is in the meteorological conditions and comparison of the SAR organizations. The third research question, where the suitability of the techniques is briefly evaluated, is discussed in section 6.4 which summarizes the main findings.

5.1 Comparison of SAR organizations

5.1.1 SAR missions in relation to meteorological conditions

This sub-section is focusing on the first hypotheses: “Wind speed and wave hgiht are high during a bad weather”. The parameters *WindSpeed* and *WaveHeight* are selected for the analysis, because those parameters are considered to have an impact on boating. First the weather parameters were investigated with the Multiform Scatterpolot Matrix separately for every organization. The patterns of the weather data were all similar for every organization. En example of a Multiform Bivariate Matrix representation of the comparison of weather parameters is found in Figure 13 in Appendix 6 (9/9). Therefore, we are more interested on those parameters that are assumed to have a significant impact on boating. The average wind speed (m/s) and the significant wave height (m) are selected for this part of the analysis.

The wind speeds are classified according to the wind warnings prepared by the FMI. The highest measured wind speed in this data set is less than 17 m/s. Wind speeds can thus be categorized into these two warning limits:

- warning for heavy wind: 10-minute average wind speed 14–20 m/s ,
- strong wind advisory: 10-minute average wind speed 11–13 m/s or strong gusts (only during the summer season, May −October).

The FMI also provides warnings on wave heights which refer to the significant wave height. The used risk levels are the following:

- 2.5 meters,
- 4 meters, and
- 7 meters.

All significant wave heights in this dataset are less than 4 meters. The warning of 2.5 meters is only considered in this analysis. In the Baltic Sea, when the significant wave height reaches this limit, sea traffic may be disrupted. The highest waves that reach the highest risk level are occurring only in the Northern Baltic Sea and southeastern part of Sea of Bothnia, which are not parts of this research area. But already a 1-meter wave height makes boating more difficult. In the reality, individual waves may be one and a half times higher than the number reported in the warning. (Finnish Meteorological Institute 2014.)

It should be noticed that all the significant wave heights derive from one buoy off Helsinki. This was mentioned in chapter 4. Wave heights only from that buoy are given to the missions, despite of the actual location of the incident.

The next images are presenting the incident types (1: *Rescue*, 2: *Assistance*, 3: *Ambulance and emergency response*, 4: *Operation suspended*, 5: *Search*, 6.1−6.8: other, see Appendix 2), SAR organizations by which the missions are performed (1: the Finnish Rescue Services, 2: the Finnish Lifeboat Institution, 3: the Finnish Border Guard), the received alarm time (0.0−1.0), day number of the week (1−7), and the month (1−12). The attribute of *AlarmTime* indicates values between 0 and 1, where 0 indicates the start of a new day. Number 1 indicates mid-night and the final end of the day. The results for the wind speeds and wave heights are separately plotted and highlighted in a PCP. Alongside, a bivariate map is used for plotting the location in the Gulf of Finland. For the combination of the two variables Multiform Bivariate Matrix is applied.

Selection of the polylines of the heavy wind warning (14−20 m/s) from the PCP plots the following results in Figure 45 and Figure 46.

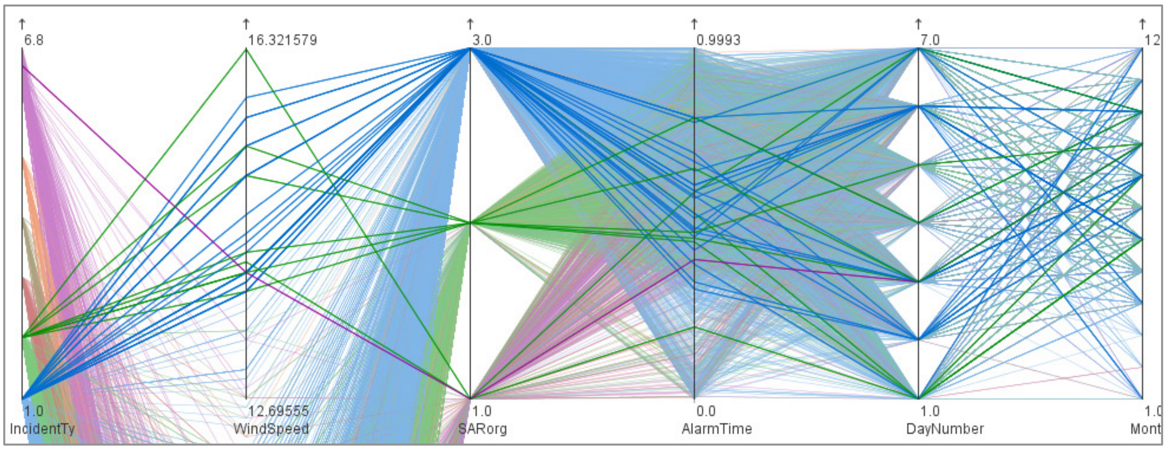


***Figure 45** PCP with highlighted polylines for heavy wind warning (14-20 m/s). The polylines were selected from the second leftmost axis (WindSpeed). The code number for attributes (from left: IncidentType, SARorg, AlarmTime, DayNumber, and Month) see Appendix 2.*

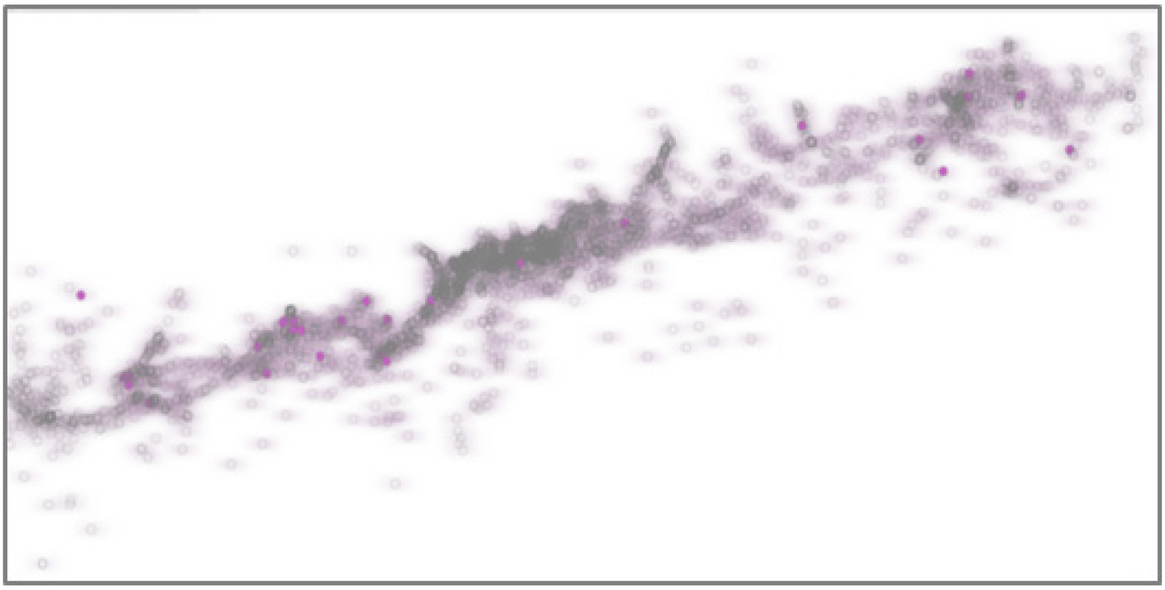


***Figure 46** A bivariate map representation of incidents highlighted in the PCP (Figure 64), when heavy wind warning (14–12 m/s) was prevailing.*

The PCP shows that the dominant incident group, when the warning of heavy wind is prevailing, is *Rescue* (blue). According to the plot, all rescue operations are performed by the FBG (3rd leftmost axis, value 3). The other incident group is *Assistance* (green), and one polyline is pointing to a minor group (purple), which is assigned as an *Environmental accident/prevention* (code numbers listed in Appendix 2). All missions performed by the FLI are *Assistance* missions. The FRS is mainly operating assistance missions and environmental accident or prevention. The alarms of the incidents were mostly received during the day time, when looking at the fourth vertical axis from the left of the PCP (AlarmTime). The second rightmost axis shows that rescue missions took place mainly during weekends or at the beginning of the week, whereas assistance missions are more evenly spread among weekdays. The rightmost axis indicates that missions during heavy wind warning were not often obtained during winter months, which was expected. The map in the lower window (Figure 46) represents the locations of the missions. The most missions took place in the western and eastern parts of the Gulf of Finland.

From these images (Figure 45 and Figure 46) it can be highlighted that the FBG was the most active SAR organization during the heavy wind warning. The most common incident type of the FBG was *Rescue*. The FLI obtained missions of type of *Assistance*, and the FRS performed the least number of missions under the heavy wind warning, mostly type of *Assistance*. The incidents took place in the western and eastern parts of the Gulf of Finland.

The next images (Figures 47 and Figure 48) show the missions that happened when the strong wind advisory (11−13 m/s) was obtained.

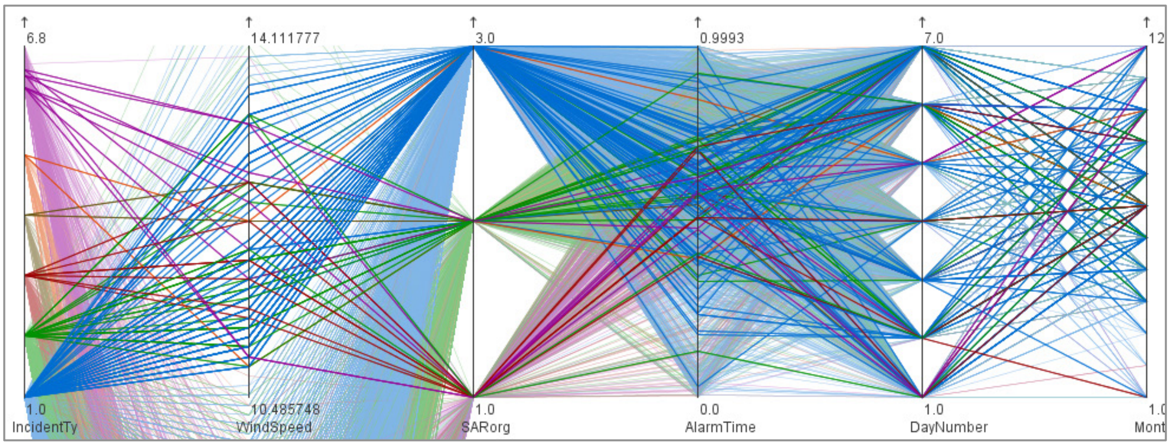


***Figure 47** Highlighted polylines for strong wind advisory (11-13 m/s). Brushing applied to the second leftmost axis (WindSpeed). Cone values of other attributes (from left: IncidentType, SARorg, AlarmTime, DayNumber, and Month) see Appendix 2.*

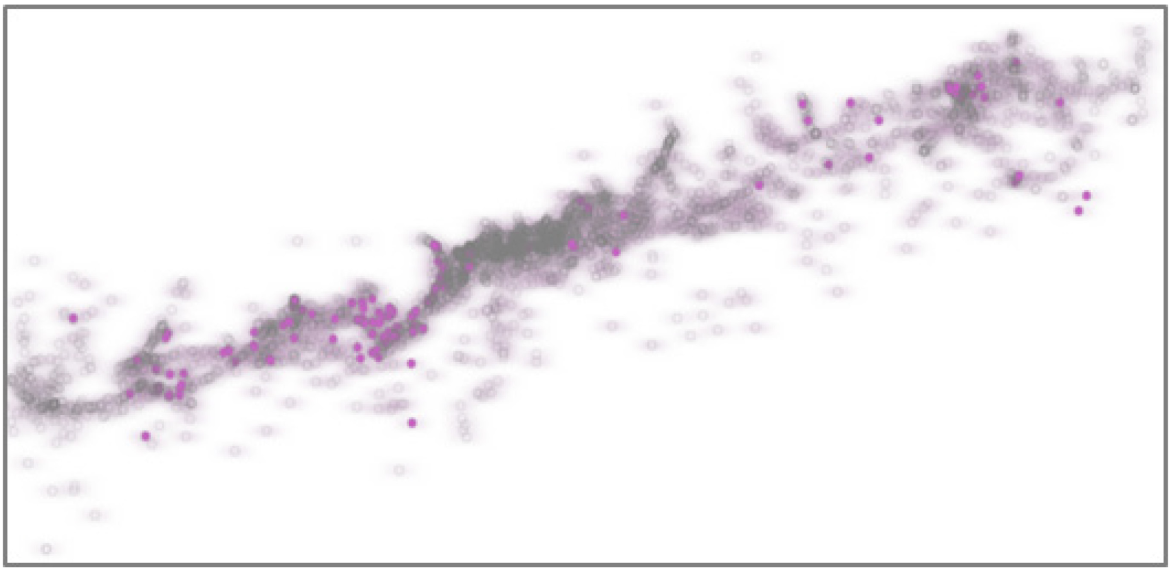


Figure 48 *A bivariate map representation on missions during a strong wind advisory.*

The pattern of the polylines and the major incident groups are similar to the previous PCP representation: *Rescue* missions are performed by the FBG, and *Assistance* missions are operated by the FLI. However, within the strong wind advisory there seem to be more other incident types involved. *Ambulance and emergency responses* (red polylines) are mainly undertaken by the FRS as well as the minor incident groups (purple polylines). Brown polylines indicate *Operation suspended*, which are operated by the FLI only. Search missions are performed either by the FBG or the FLI. Not particular differences occur in the timing of the day, shown in the fourth axis (*AlarmTime*). Alarm mostly was received during the day. Also the axes of *DayNumber* and *Month* are now more widely and evenly spread. The winter months still seem to be inactive of operations.

On the other hand, the bivariate map (Figure 48) shows that the missions are even more clustered in the western and eastern parts of the Gulf of Finland. The biggest clusters are in the West, from Raasepori to Kirkkonummi. One fact has to be noted that Hanko area was not provided with wind-related data. Therefore the missions in Hanko area are not plotted. It is still assumed that Hanko is a windy place, because it is the outermost to the open sea.

From Figures 47 and 48 the findings are similar to the findings in previous Figures 45 and 46. The FBG is the most active during these circumstances related to the wind. Certain incident types are typical for certain SAR organizations. The FBG performs *Rescue* missions, the FLI *Assistance*, and the FRS seems to have the most variety of different incident types.

The next images (Figure 49 and Figure 50) are similar PCP and bivariate map representations, but are highlighting the significant wave heights which approximately reach the minimum limit of wave warning which is 2,5 meters high or more.

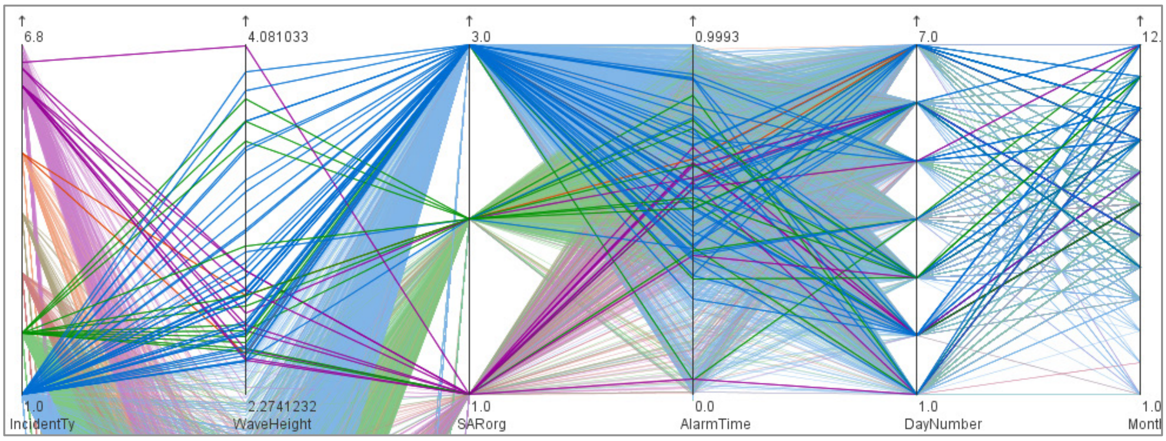


Figure 49 *PCP representation of wave heights for wave warning (2,5 m or more). Polylines brushed fron the second leftomost axis (WaveHeight). Cone values of other attributes (from left: IncidentType, SARorg, AlarmTime, DayNumber, and Month) see Appendix 2.*

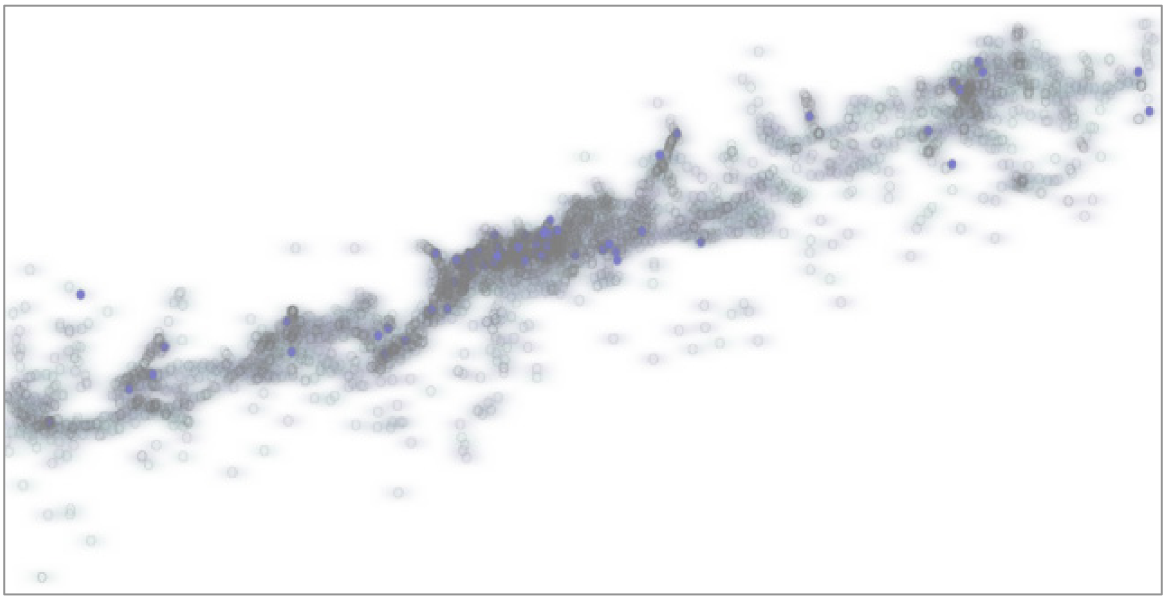


Figure 50 *Locations of the incidents with wave warning.*

The pattern of the PCP representation is similar to the wind-related data. The FBG again seems to remain the most active SAR organization with *Rescue* operations. *Assistance* seems to be the dominant incident type for the FLI. Some variety occurs in the polylines intersecting the third leftmost axis (SARorg) in value 2 (see Appendix 2). Blue, orange and purple polylines are intersecting the axis. Blue indicates *Rescue* incidents, orange *Search* incidents and purple is indicating the minority groups of incidents, which are mainly performed by the FRS. The map in Figure 50 show the locations of incidents assigned with high waves. They are mostly located around the metropolitan area. Some smaller groups can be found in the West and in the East.

It should be noticed again in this case that the wave height derive from one buoy only, which is located in the open sea off Helsinki. The real wave height near coast or in bays may be different. The wave heights are not accurate in these locations.

At this point, it is clear that the FBG is the most active Maritime SAR organization, although all SAR organizations are on duty despite the weather or time of the day. Winter months, especially January–March, are less active months for SAR missions. *Rescue* is the most common incident type for the FBG, *Assistance* for the FLI and *Ambulance and emergency responses* as well as the minor incident types (group 6, see Appendix 2) are mainly obtained by the FRS.

The order of the axes in PCP is important if we want to look at the relationship between two variables. The axes of those variables need to be placed parallel next to each other. Figure 51 shows an example, how vertical the polylines are between *WaveHeight* (middle axis) and *WindSpeed* (rightmost axis). If the polylines are vertical it means that there is a positive relationship between the variables. There seem to be strong relationships especially among the lower wave heights and wind speeds. In a huge dataset, the clutter problem is occurring in this case as well. The stack of polylines is full of blue and green polylines (because those are the dominant groups), which make other colors harder to distinguish. It is also not possible to select combinations of two variables with a PCP technique.

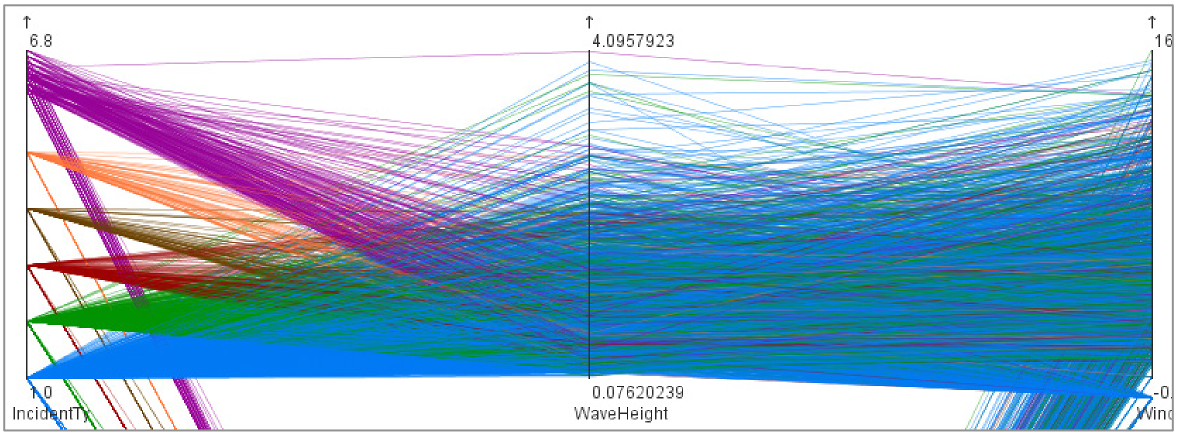


Figure 51 *The polylines are hard to distinguish due to cluttering.*

In order to select those incidents with both high significant wave height and high average wind speeds other technique has to be used. In this case, a scatterplot is suitable for this purpose. The next figure of windows (Figure 52) represents detailed scatterplots selected from the Multiform Bivariate Matrix. In the first window from the left, there are selected those points that have the minimum risk level of heavy wind warning (around 4 m/s) or more and the minimum risk level of wave warning (around 2,5 m) or more. Wind speeds are on the horizontal axis and wave heights on the vertical axis. The middle window shows which SAR organization (horizontal axis, values 1.5 and 2.5 are meaningless in this case) has performed what types of incidents (vertical axis) under high wave and wind conditions. The rightmost window correspondingly explains the municipalities (horizontal axis, 0–15, see Appendix 2) where the incident types occurred.

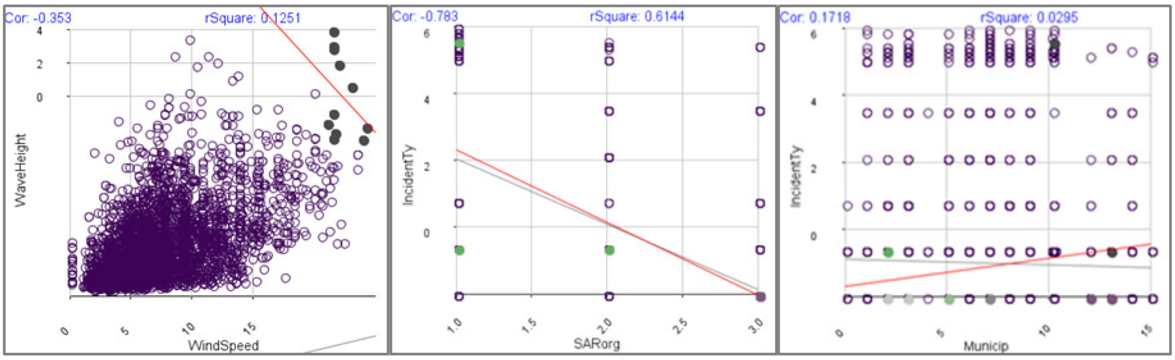


Figure 52 *Detailed scatterplots of wind speeds and wave heights (left), SAR organizations and incident types (middle), and municipalities and incident types (right). The value codes of attributes provided in Appendix 2.*

It is shown that not many incidents share the highest values of the both attributes. They seem also to be located in the western and eastern parts of the gulf, but do not show how many. The bivariate map (Figure 53) plots those incidents as brown dots. As noted, there are not many incidents. The incidents are very sparsely located along the coast of the Gulf of Finland.

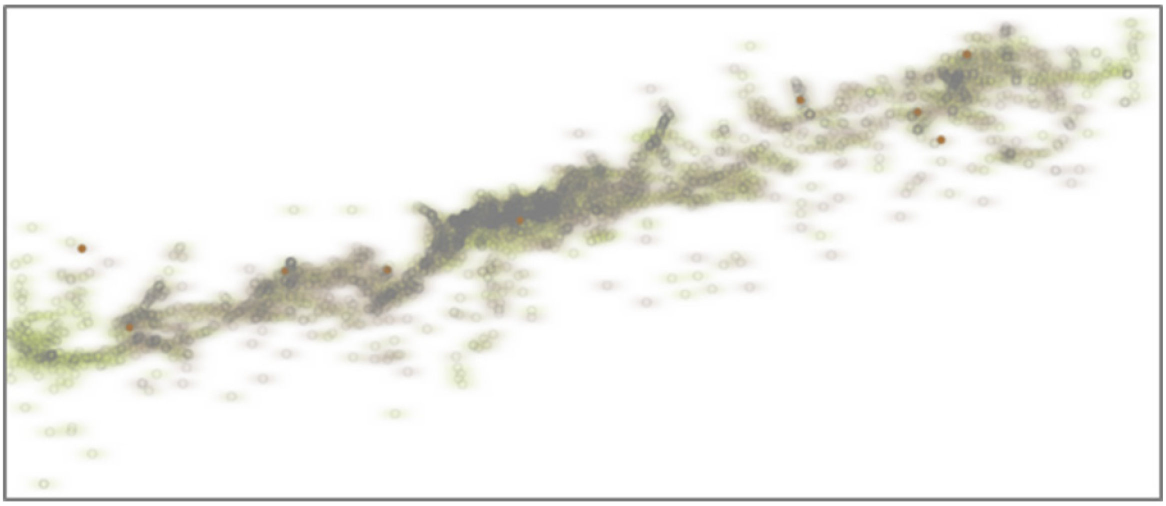


Figure 53 *Bivariate map representation of incidents reaching the risk limits of heavy wind warning and wave warning.*

This small amount of incidents does not give further conclusions. The next images (Figure 54 and Figure 55) show the same applied technique for the same variables, but with the minimum risk limits of strong wind advisory (around 11 m/s) or more and 1-meter wave height or more.

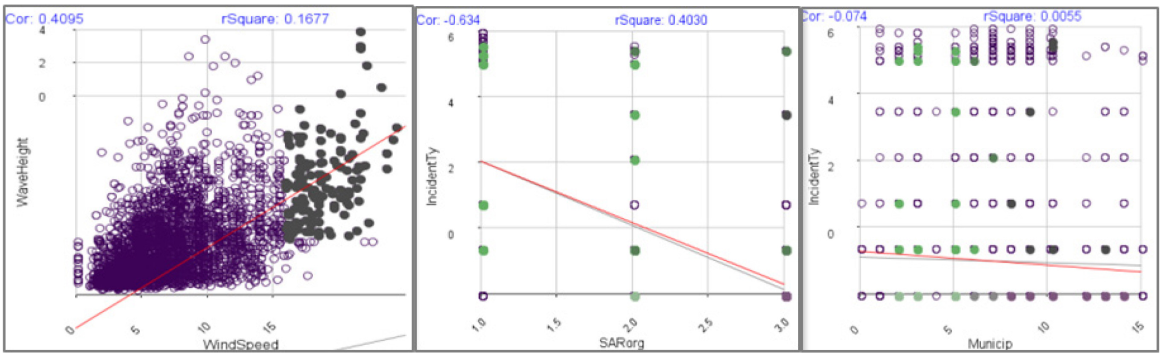


***Figure 54** Scatterplots presenting selected items that reach the strong wind advisory and 1-meter wave warning (leftmost window).*

Looking at the plots, more incident types come involved for every organization. The rightmost window shows the municipalities for the occurrence of incident types. The map (Figure 55) plots clearly the positions of the incidents. It seems that the western part of the gulf is the dominant area of incident in challenging weather conditions.

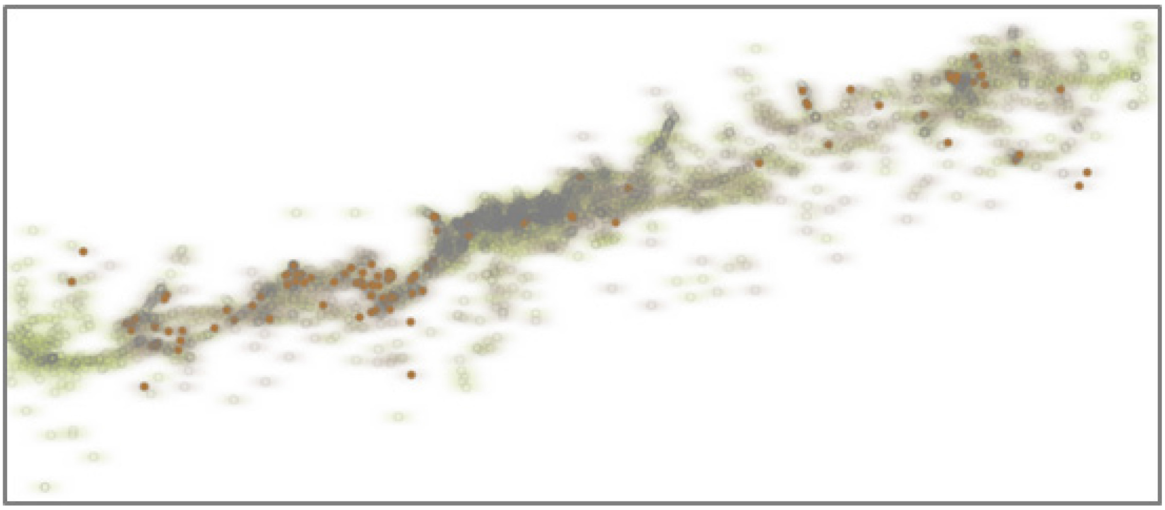


***Figure 55** Plotted incidents with challenging wave and wind conditions.*

After having investigated the performances during challenging conditions, the FBG seem to operate missions the most actively. The majority of the missions are *Rescue*-oriented, and *Assistance*-oriented missions are operated mostly by the FLI. The FRS carries out the mission types that are common on land such as wild fires, building fires, fire fightings and loss or environmental accidents or preventions. *Ambulance and emergengy resonses* were responded mainly by the FRS during bad weather. Wind conditions seem to be challenging in the western and far in the eastern Gulf of Finland, as well as waves, although the highest waves were located around metropolitan region, not forgetting that the wave heights indicate waves in the open sea not the coastal sea. During the summer, it is common that tourist travel to the Finnish archipelago, which might explain why, especially, the western part of the gulf is cluttered the most.

5.1.2 Temporal and spatial analysis

In this section, the aim was to compare the different locations of the Gulf of Finland as well as the incident occurrence during different times of the day visually mainly with PCP and bivariate map. The focus of this sub-section is the second hypothesis introduced at the beginning of this chapeter: “Most missions are performed during the day hours. What is the pattern during the night and late-evening hours?” It is obvious that incidents occur mainly during the day time and early evening. Due to the bright summers, mid-summer party, attraction of archipelago and other events, people are spending more time in seas around the day.

This section investigates the incident occurrence in general, during night hours (around 0:00 a.m – 5.00 a.m.) and late-evening hours (around 9.30 p.m. – 0:00 a.m.). First, the comparison is made for the whole dataset including incidents of all SAR organizations. After that, the comparison is obtained for the incidents of the FLI.

The area is divided into three parts: the western Gulf of Finland, the metropolitan area, and the eastern Gulf of Finland. The western part of the gulf consists of four leftmost municipalities, Hanko, Raasepori, Inkoo, and Siuntio. Then follows the metropolitan area, in which the next five municipalities are included: Kirkkonummi, Espoo, Helsinki, Sipoo, and Porvoo. The eastern part includes the five rightmost municipalities: Loviisa, Pyhtää, Kotka, Hamina and Virolahti (Figure 56). Proportions of each incident type have been calculated for the entire dataset and for each area. The proportion values are compiled in a table attachen in Appendix 5 (1/2). The other similar table in Appendix 5 (2/2) shows the ratios of the FLI incidents.

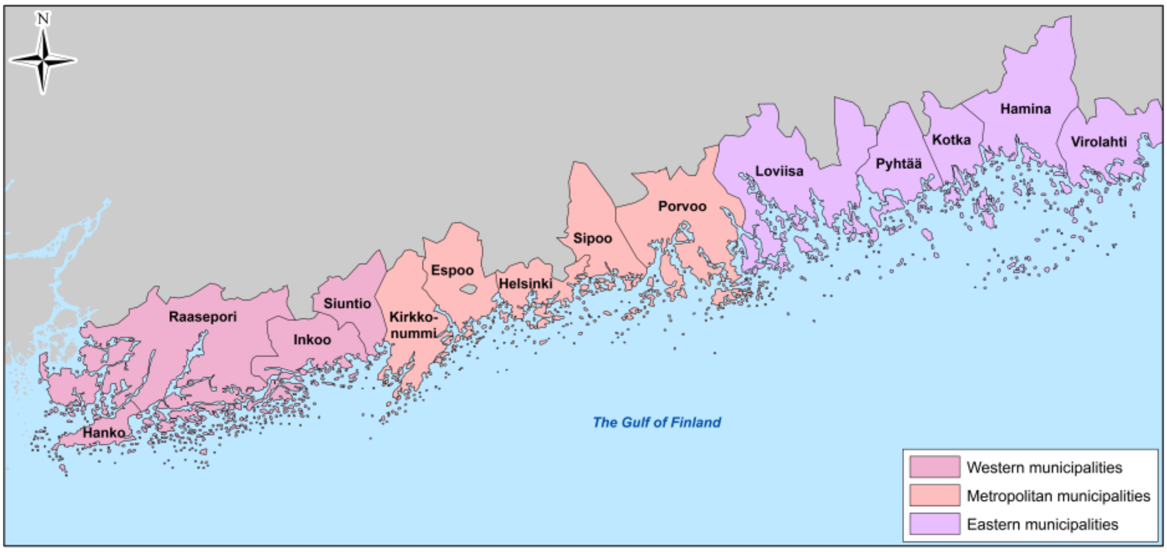


Figure 56 *Map representing the division of the coast into three areas.*

The values in the Appendix 5 (1/2) show that the metropolitan area is the most active region in SAR operations. Around 70 % of the missions have been undertaken in that area. When comparing the proportions of the three major incident classes in a general level (*Rescue*, *Assistance*, *Ambulance and emergency response*), there are varying values between the areas. The highest proportion of *Rescue* incidents is in the eastern part of the gulf. Metropolitan area has the highest proportion of *Assistance* incidents, and *Ambulance and emergency responses* in the West. The main differences of the proportions occur among these three major classes. There are no particular differences during the night and evening hours either, because the missions are perfomed mainly during day hours. Except one difference in the western part during late-evening hours is identifed in the proportion of the three major classes together. When comparing at the incident classes separately, the proportions of *Rescue* and *Ambulance and emergency responses* are higher in the evening than at night. Proportions of the *Rescue* missions is higher in the evening than at night in every region (around 4 – 5 %). This can be explained with the fact that the evenings in summer are bright and the archipelago, especially in the West, is attracting tourists.

Since the main differences occur among the three classes, we shall select those classes and explore them visually in these regions. Figure 57 shows three snapshots of bivariate map representations of the incidents in the three areas in general. The lightest green color indicate *Rescue* incidents, *Assistance* incidents is the darker blue-green color, and the darkest blue-shaded hue shows the *Ambulance and emergency responses*.

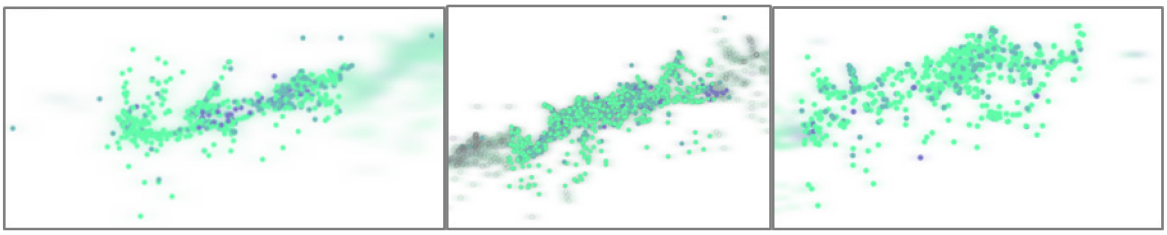


Figure 57 *Three major incident classes plotted for every region; light green: Rescue, darker green: Assistance, blue: Ambulance and emergency response.*

Some of the incidents seem to reach far from the main land. However, the coast line is very tattered with lots of islands which are not visible in the maps. The incidents are not necessarily located far from the shore line. The can be near an island.

In the next figures, PCP technique is used with the bivariate map. In the PCP, the applied variables (from the left) are *IncidentType* (containing only the three major incident classes), *DayNumber*, *Month*, *Municipality*, *SAR organizationa* and *AlarmTime* (see value codes in Appendix 2). The relevant variables to be compared are the three last ones, because the bivariate map shows the locations of the incidents and PCP is used for the comparison of the SAR organizations. Blue polylines are *Rescue* operations, green polylines are *Assistance* missions and orange polylines indicate *Ambulance and emergency responses*.

The next figures (Figures 58 and 59) present the western Gulf of Finland during night (upper window) and evening hours (lower window). The hours are selected by brushing the values of the axis of *AlarmTime* in the PCP (rightmost axis). The axis of municipalities (third rightmost) is zoomed to show the municipalities of the incidents on the map.

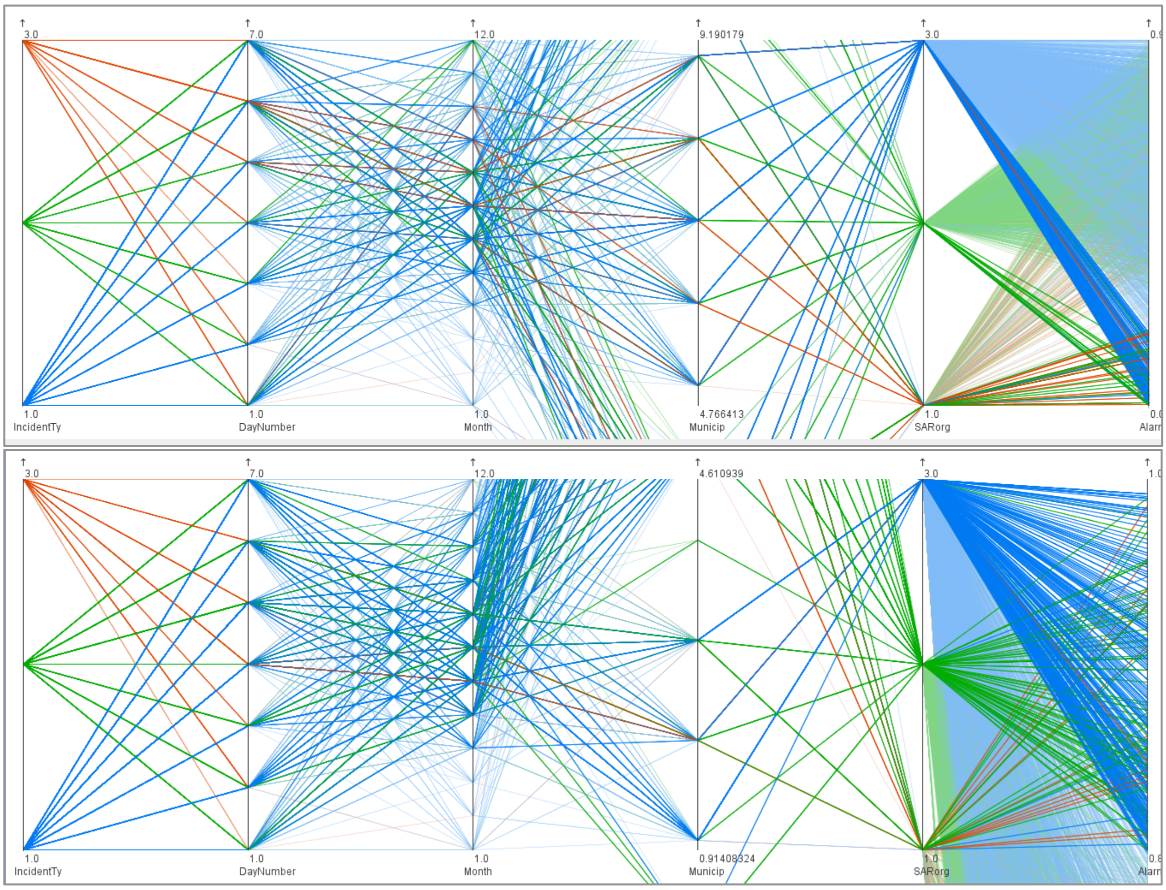


Figure 58 *Rescue, Assistance, and Ambulance and emergency operations in the western Gulf of Finland during the night (upper window) and evening hours (lower window). Attributes from the left: IncidentType, DayNumber, Month, Municipality, SARorg, and AlarmTime. Code values provided in Appendix 2.*

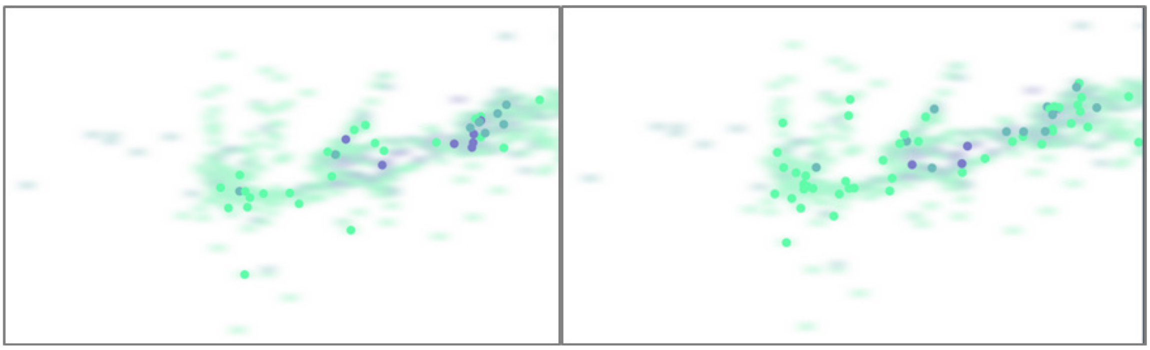


Figure 59 Western municipalities during the night (left) and the evening (right; light green: Rescue, darker green: Assistance, blue: Ambulance and emergency response).

The incident on the map images are located near to the shore line compared to the general view (Figure 57). *Rescue* operations occur during the evening hours more often than during night hours. Number of *Assistance* missions are about the same in both figures. *Ambulance and emergency responses* are not as common as the two other types of incidents, but seem that they have been operated during the night.

The PCP distinguishes the incident types between the SAR organizations: *Rescue* missions are mainly performed by the FBG, the FLI obtained the incidents in a need of assistance, and the *Ambulance and emergency responses* are mainly addressed for the FRS, who also has given some assistance. This is a known pattern that appeared from the previous findings concerning wind and wave conditions.

Figure 60 and Figure 61 present the same techniques on incidents in the metropolitan area.

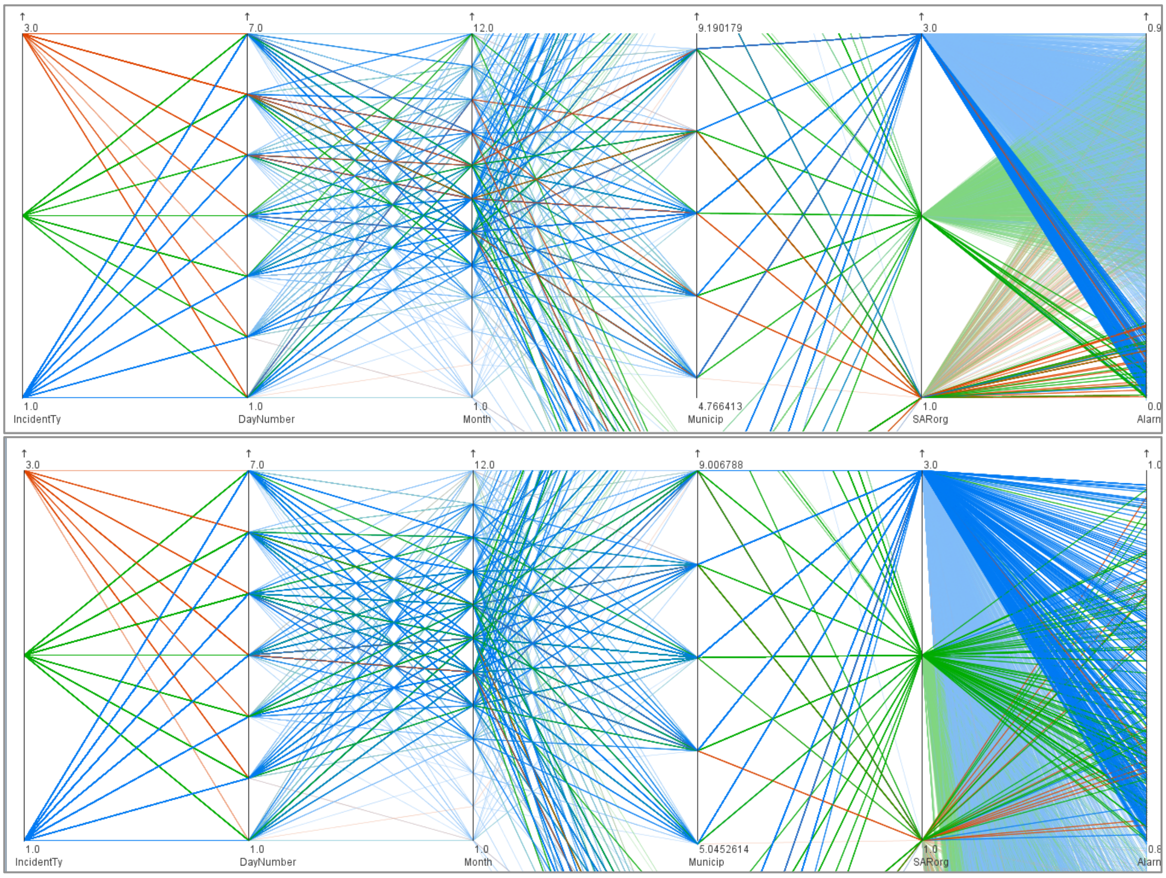


Figure 60 *PCPs plotting incidents in the metropolitan area during night (upper window) and evening (lower window). Attributes from the left: IncidentType, DayNumber, Month, Municipality, SARorg, and AlarmTime. Code values provided in Appendix 2.*

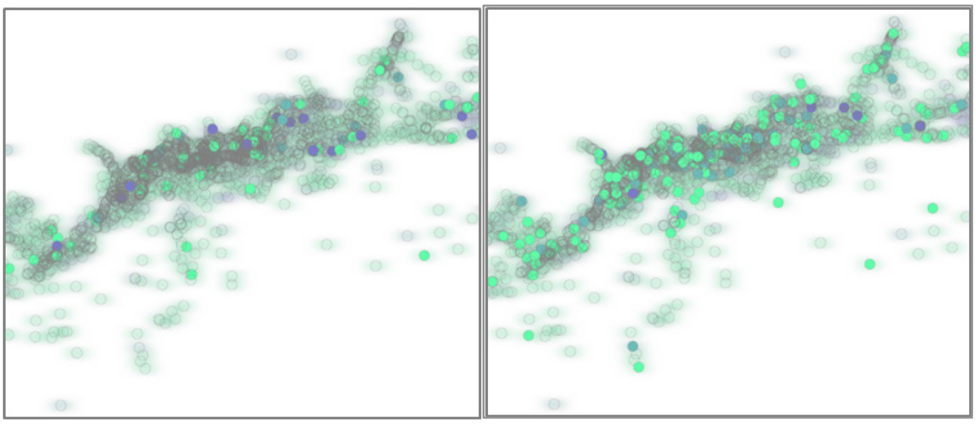


Figure 61 *Incidents in the metropolitan area during the night (left) and evening (right); light green: Rescue, darker green: Assistance, blue: Ambulance and emergency response.*

The incidents in Helsinki region are very densely located, which indicates the activity of the region in SAR operations. The clusters may hide some incidents or make the incidents hard to distinguish. Generally, the biggest incident class is formed by *Rescue* missions, but incidents of *Assictance* has the biggest proportion in this region (see Appendix 5 (1/2)). The pattern is again similar to the western region, except the incidents are more cluttered to the shore. Again, the incidents are closer to the shore (or to an island) during the night and late-evening hours. The occurrence of the incidents are following the same trend: there have been more *Rescue* operations in the evening than at night, *Ambulance and emergency responses* occurred more often during the night, and no difference in *Assistance* operations. The PCP (Figure 59) shows the same division of missions between the SAR organizations: *Rescue* obtained mostly by the FBG, *Assistance* provide the FLI, and *Ambulance and emergency responses* are undertaken by the FRS.

Comparing the missions in the eastern Gulf of Finalnd (Figure 62 and Figure 63) is also not givin any more expetional results. The trends are the same in the East as well. Each SAR organization carries out one of these missions more specifically than another SAR organization: *Rescue* – the FBG, *Assistance* – the FLI, and *Ambulance and emergency response* – the FRS. The main differences are shown in the proportions calculated for each region (Appendix 5 (1/2)). The East has the highest ratio of *Rescue* missions, the metropolitan has the highest ratio of *Assistance*, and the highest proportion of *Ambulance and emergency responses* are in the West. *Ambulance and emergency responses* seem to have been very rare in the East both during night and the evening.

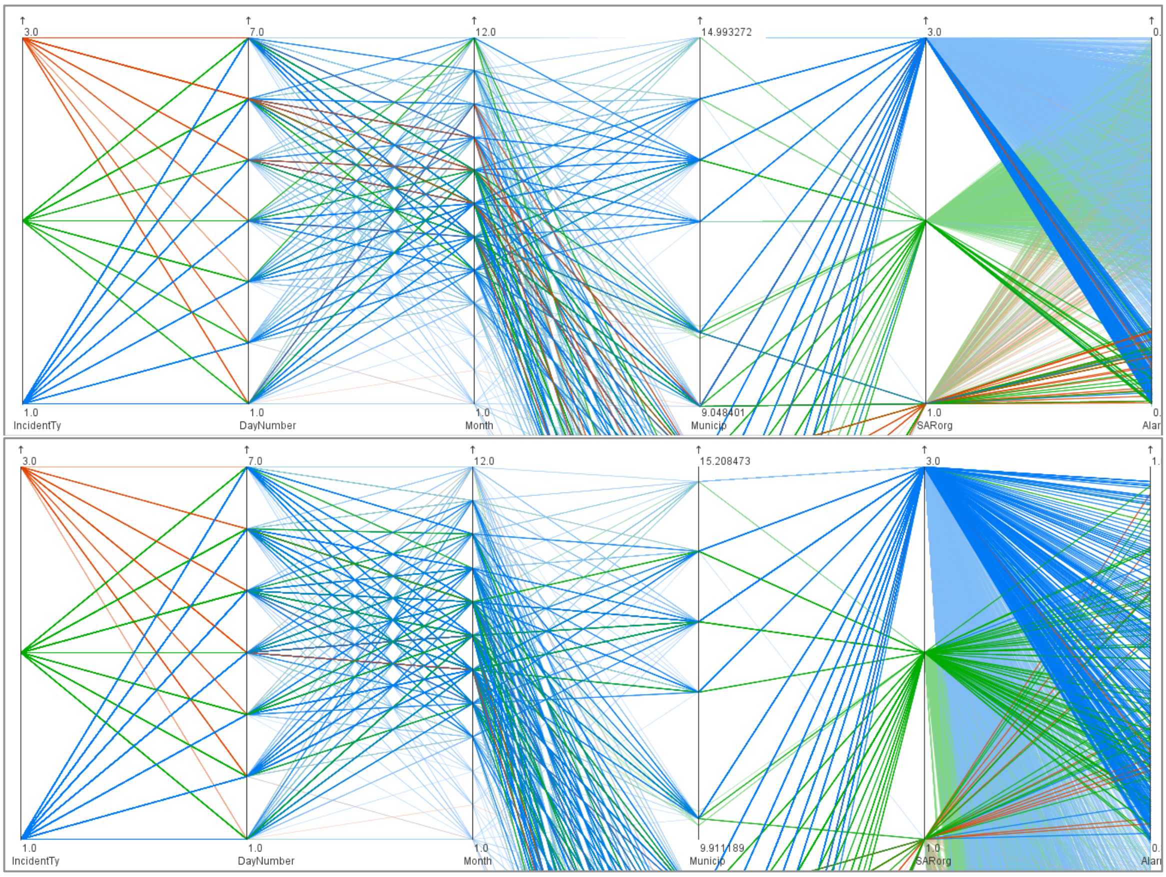


Figure 62 *The eastern Gulf of Finland with rescue, assistance and ambulance/emergency missions. Night hours (upper window), evening hours (lower window). Attributes from the left: IncidentType, DayNumber, Month, Municipality, SARorg, and AlarmTime. Code values provided in Appendix 2.*

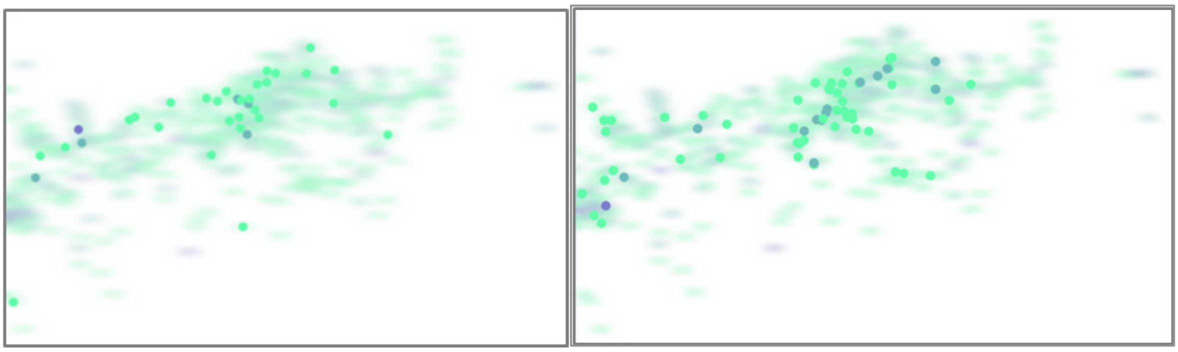


Figure 63 *Incident located during different times, night (left) and evening (right).*

Now, the three major classes have been investigated for the SAR organizations in different regions. Table in Appendix 5 (2/2) lists the proportions of the FLI incidents. The same regions, night hours and evenign hours are being used. The dominant region of incidents remains the metropolitan area, which is nearly the same ratio as in the previous table in Appendix 5 (1/2).

It was identified that the major incident class is *Assistance* for the FLI. The *Rescue* group has the highest ratio in the West and not in the East as it is in the other table in Appendix 5 (1/2). The highest proportion of *Assistance* is in the eastern Gulf of Finland. *Ambulance and emergency response* is not one of the most common incident types for the FLI, which is seen from the proportion values. *Operation suspended* is only perfomed by the FLI, and this class in the third dominant incident group of the FLI, which is has the hihgest ratio in the metropolitan region. The other incident types performed by the FLI include *Loss prevention*, *Firefighting*, *Environmental accident/prevention* and *Other mission*. There are not significant differences between the regions during different times of the day.

The PCPs in Figure 64 briefly describe the incidents of the FLI during night and late-evening hours. The proportion values of the missions are not high during these times of the day. *Rescue* (blue) and *Assistance* (green) missions are not common. Some polylines of *Search* class (red) are highlighted in both PCPs. The FLI has performed suspended operations (orange) during the late-evening hours (lower window). The polylines of minor incident classes (purple) have been operated during the night time (upper window).

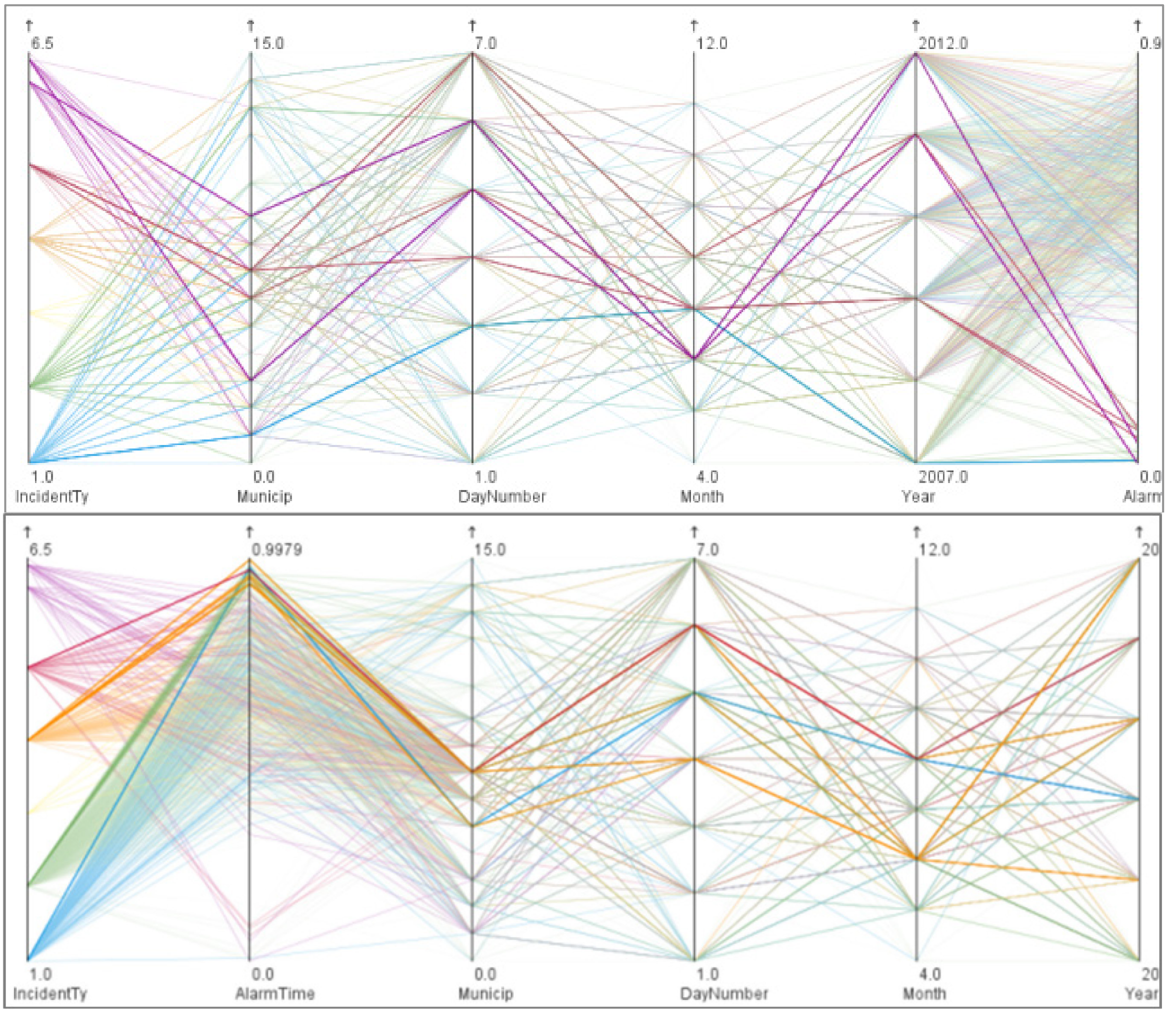


Figure 64 *FLI mission during the night hours (upper window) and late-evening hours (lower window). Attributes shown in PCP: IncidentType, AlarmTime, Municipality, DayNumber, Month and Year. Code values provided in Appendix 2.*

The evening incidents seem to have located in the metropolitan region mostly, whereas the night incidents occurred also in the West and some in the East when looking at the Municipality axis in both plots. The night cases occur during summer months (June – August) and mainly at the end of the week. The cases in the evening mainly happen during weekends (Friday – Saturday) of summer months (July and August).

5.1.3 Travel distances

The SAR resources and the environment are always limited. Distribution of efficient and successful SAR response may be a challenging task due to limitations of time, distance, weather and sea conditions, currents, and visibility. In urgent situation, time can be very crucial when aid is needed. The alarm has to be responded within a certain time limit. The equipment needs to be adequate for the occasion. Depending on the situation, certain kinds of equipment are suitable for certain kinds of conditions. The aim is to deliver SAR service to the destination. The most obvious barrier between the SAR resource and the target destination is the distance. In addition, other conditions such as environmental conditions might cause friction to the delivery, although the SAR equipment may reach the destination. With a fast response and adequate SAR equipment such as vessels and other resources are widening the radius of SAR response.

This section compares the traveled distances of the SAR organizations focusing on the last hypotheses: "The radius of action is varying between the SAR organizations, meaning there are differences in travel distance between the missions and the SAR station and between the missions and the shore." The distances are compared between the incidents and SAR stations or between incidents and shore. It has to be remarked that shoals and low grounds are not taken into account and the available information on SAR stations is varying between the organizations.

Distances were not provided with the original data. The distances were calculated by using ArcGIS 10.1 and may thus not represent the real distances. The data of the FRS was not provided with station information. The FBG has eight stations, but the incidents are not provided with the information, which station has undertaken the incident. Only the FLI provides the incidents with the performing station.

The applied techniques are scatterplot and PCP. The scatterplot is mainly used for selecting the distances and possible combinations. The purpose of PCP is to show incidents and compare other variables.

The next Figure 65 represents a scatterplot (upper window) and a PCP (lower window) representation of the selected items of the FBG. The scatterplot is comparing the distances between the incidents and the shore (horizontal axis) and between the incident and the closest SAR station of the FBG (vertical axis). The majority of the incidents are not very far away from the shore or from the station. Distances of more than 10 km from both the shore line and SAR station are rare for the FBG. The selected items in the scatterplot represent those incidents that reach or exceed 10 km distance. The PCP highlights the polylines of the selected items. According to the view, polylines of *Rescue* operations (blue) and some *Assistance* operations (green) reach this distance. The second axis from the left indicates the closest SAR station of the FBG. Assumedly, the incidents are located quite evenly along the gulf, except far away in the West and in the East, since Hanko, Haapasaari and Hurppu station do not have any highlighted intersecting polylines (seen on the second leftmost axis, CloseRVLst).

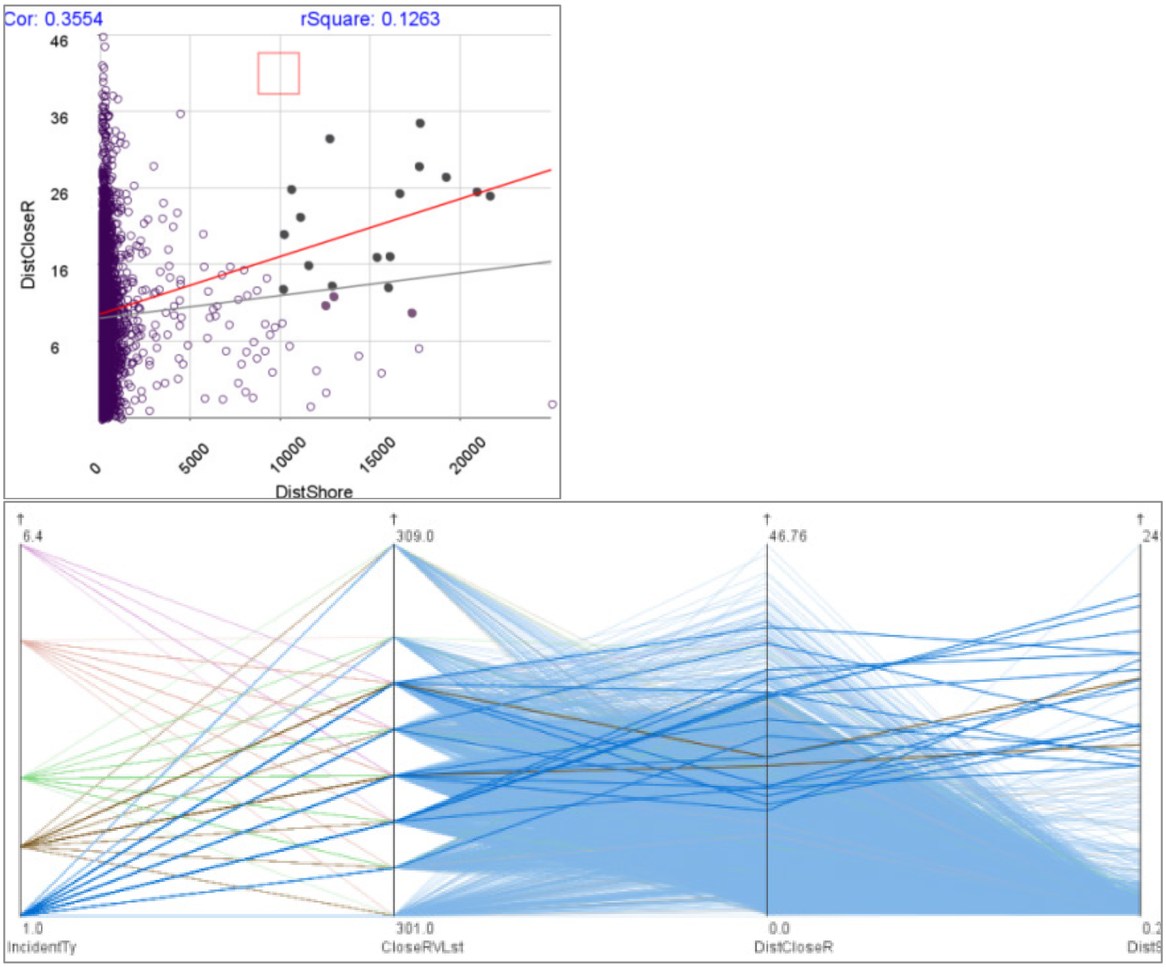


Figure 65 *Mission of the FBG with long travel distance, 10 km or more from the shore and from the closest SAR station. Scatterplot (upper window) is suitable for detecting long distances from the shore (horizontal axis) and from the SAR station (vertical axis). PCP (lower window) highlights the selected items of the scatterplot. See value codes for the attributes IncidentType and SAR stations of the FBG in Appendix 2.*

The next two Figures 66 and 67 compare the distances calculated for the FLI using the same techniques as in the previous example (Figure 65). The selected items of the scatterplot in Figure 66 are the incidents that are more than 3 km away from the shore, because it does not seem to be very common for the FLI. Polylines of *Assistance* and *Rescue* operations (leftmost axis) are again highlighted for the stations far in the West and the East when looking at the second leftmost axis (*ClosestSMPS*). Polylines are intersecting stations in Hanko (201), Porkkala (205), Espoo (206), Helsinki (208), Kotka (211) and Hamina (2012) (value codes for the stations see Appendix 2). The metropolitan area was identified as the busiest in sea traffic. Hanko is open to the sea and in the East heavy traffic occurs as well. These areas are popular for tourists, which is taken into account.

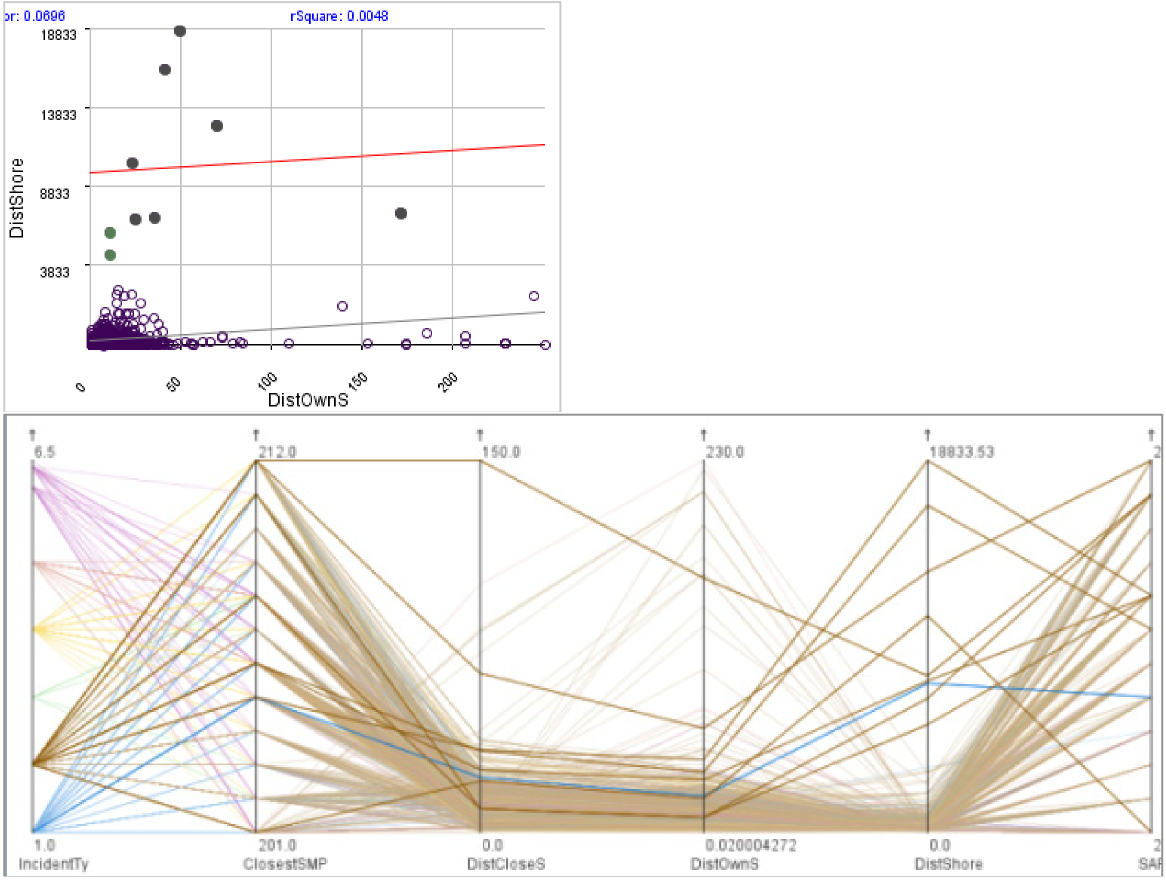


Figure 66 *Highlighted items of the FLI missions that are more than 3 km away from shore. Selection applied in scatterplot (upper window. PCP (lower window) highlights the polylines of the selection. Attribute from the left: IncidentType, ClosestSMPS (closest SAR station), DistClosesS (distance to the closest station), DistOwnS (distance to the own station), DistShore (distance to shore), and SARstation. Value codes of attributes provided in Appendix 2.*

The scatterplot of Figure 67 shows that the distance between the incident and the performing station is mostly less than 50 km. Those incidents that are located 50 km or more from the SAR station are selected. These incidents, however, are not necessarily far away from the shore. This is shown in the PCP window (lower window) on the second rightmost axis (*DistShore*). Most of the highlighted polylines are intersecting the lowest values of the axis.

Although the distance to the own SAR station is long, the distance to the nearest SAR station may be short. This can be identified from the third leftmost axis (*DistCloseSMPS*). SAR vessels of Helsinki stations seem to travel the farthest distance, which is shown in the scatterplot. In the West, Hanko and Bågaskär stations do the farthest missions and Kotka and Hamina stations in the East. The most operations are for assistance purpose (brown polylines). A couple of highlighted polylines of Search (red polylines) and a minor mission type (purple polyline) appear mainly in the West.

Some SAR stations of the FLI perform missions far away from the station, even though another SAR station could be closer to the incident. Secondly, the incident is not necessarily far away from the shore although it is from the station. The explanation for this is unclear. It might be that those SRUs may be there for a rehearsal or for a visit.

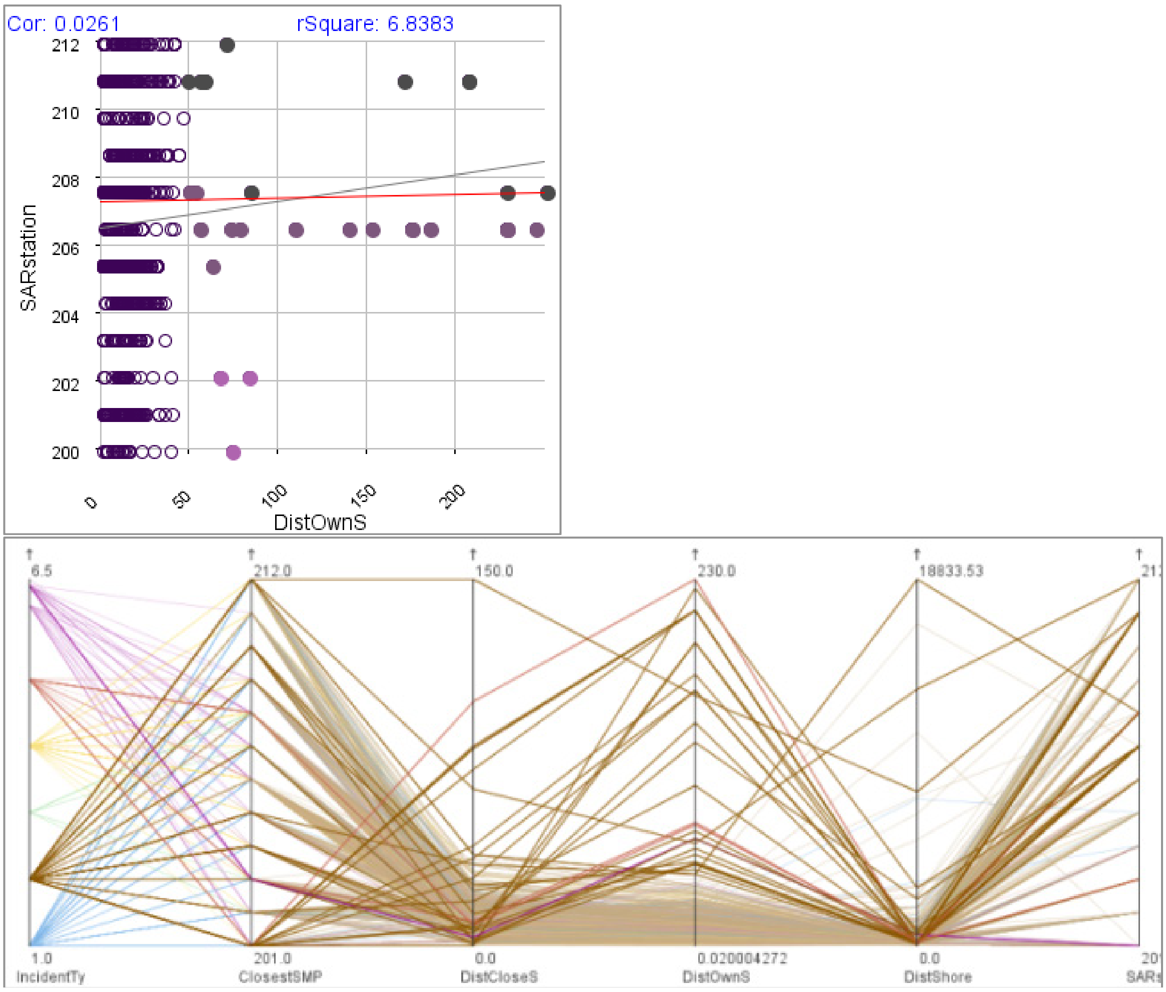


Figure 67 *Highlighted items of the FLI missions that are more than 50 km from the station. Selection applied in scatterplot (upper window). Attribute from the left: IncidentType, ClosestSMPS (closest SAR station), DistClosesS (distance to the closest station), DistOwnS (distance to the own station), DistShore (distance to shore), and SARstation. Value codes of attributes provided in Appendix 2.*

Figure 68 is a PCP representation of the farthest distances of the FRS. Since the FRS has no information related to stations, the selected distances are the farthest distances from shore. Since the maritime SAR incidents are addressed mainly to the FBG and to the FLI, the most maritime SAR missions operated by the FRS are less than a half of a kilometer away from shore (less than 400 m). The farthest cases seem to be *Assistance* (brown) and *Rescue* (blue) types. The other highlighted polylines (purple), which belong to the minor group, are common for the FRS. The missions have been operated near metropolitan area, since the polylines mainly intersect the values of SAR stations of the FBG and FLI in that area (axes *CloseSMPSs* and *CloseRVLs*, code numbers for the SAR stations see Appendix 2).

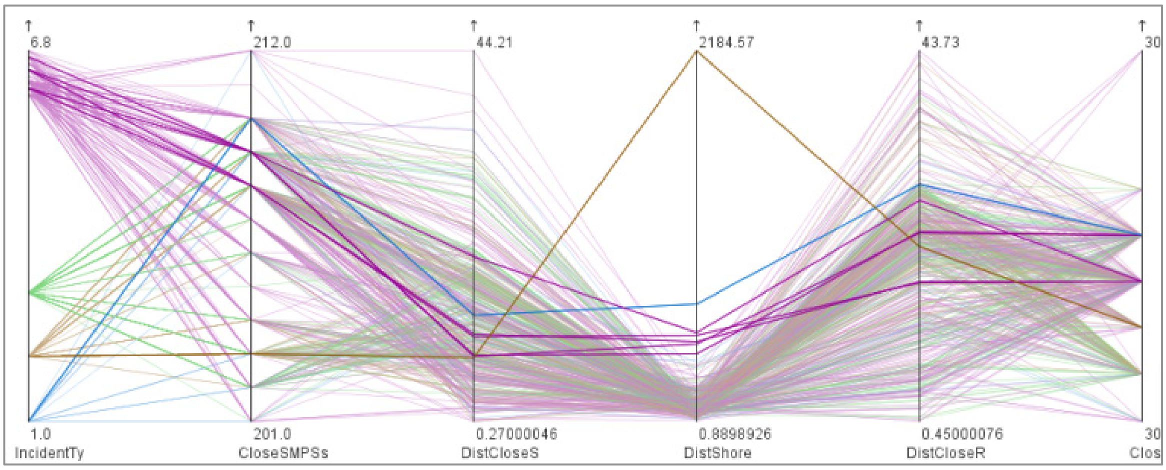


Figure 68 *The highlighted polylines of the FRS indicate cases that reach the distance of 400 m or more from shore. Attributes from the left: IncidentType, CloseSMPSs (closest station of the FLI), CistCloseS (distance to the closest FLI station), DistShore (dicstance to shore), DistCloseR (distance to the closest FBG station), and ClosestRVLs (closest station of the FBG). Value codes for the attributes provided in Appendix 2.*

The exploration of this section provided the information on travel distances of the SAR organizations. The largest radius of action has the FBG. The FLI still may operate long-distance travels, but not as far away in the open sea as the FBG. The FRS undertakes missions nearby the shore and long distances are not responded by the FRS.

5.2 Accident factors of the Finnish Lifeboat Institution

This section concentrates on the cases performed by the FLI. The exploration is part of the more detailed analysis focused on the other data that is not included in the general-level analysis such as accident factors and accident qualities as well as boat classes and boat sizes. The exploration mainly investigates the factors of the accidents which define the accident quality. The accident qualities are classified into four classes: *Unknown* (0), *Accident/Emergency* (100), *Technical* (200), and *Other* (300) (see Appendix 3). For example, the factors starting, for instance, with number 1are classified as part of *Accidnet/Emergency* class. The boat classes are defined as *sailboats*, *motorboats* or *other* that are used for recreational purpose. Boat sizes are only known for the motorboats: less than 7 m (1), 7 – 15 m (2), and more than 15 m (3). The code numbers of the boat classes and sizes are listed in Appendix 2.

The interesting finding so far is the incidents that reached the both risk limits of the strong wind advisory (11 m/s or more) and wave warning (2,5 m or more). Figure 69 represents the incidents during the strong wind advisory (11–13 m/s) and wave warning (2,5 m), but the colors in the map window (lower window) are indicating the accident qualities: blue – *Unknown*, green – *Accident/Emergency*, orange – *Technical*, and purple – *Other*. Cases assigned as *Other* are not occurring in the plots. In the East, in Kotka region, no technical accidents are identified. The accidents due to technical failures are mostly located in the West, Inkoo – Kirkkonummi area. Cases of emergency are mainly located in the same area among technical accident, but also nearby Tammisaari.

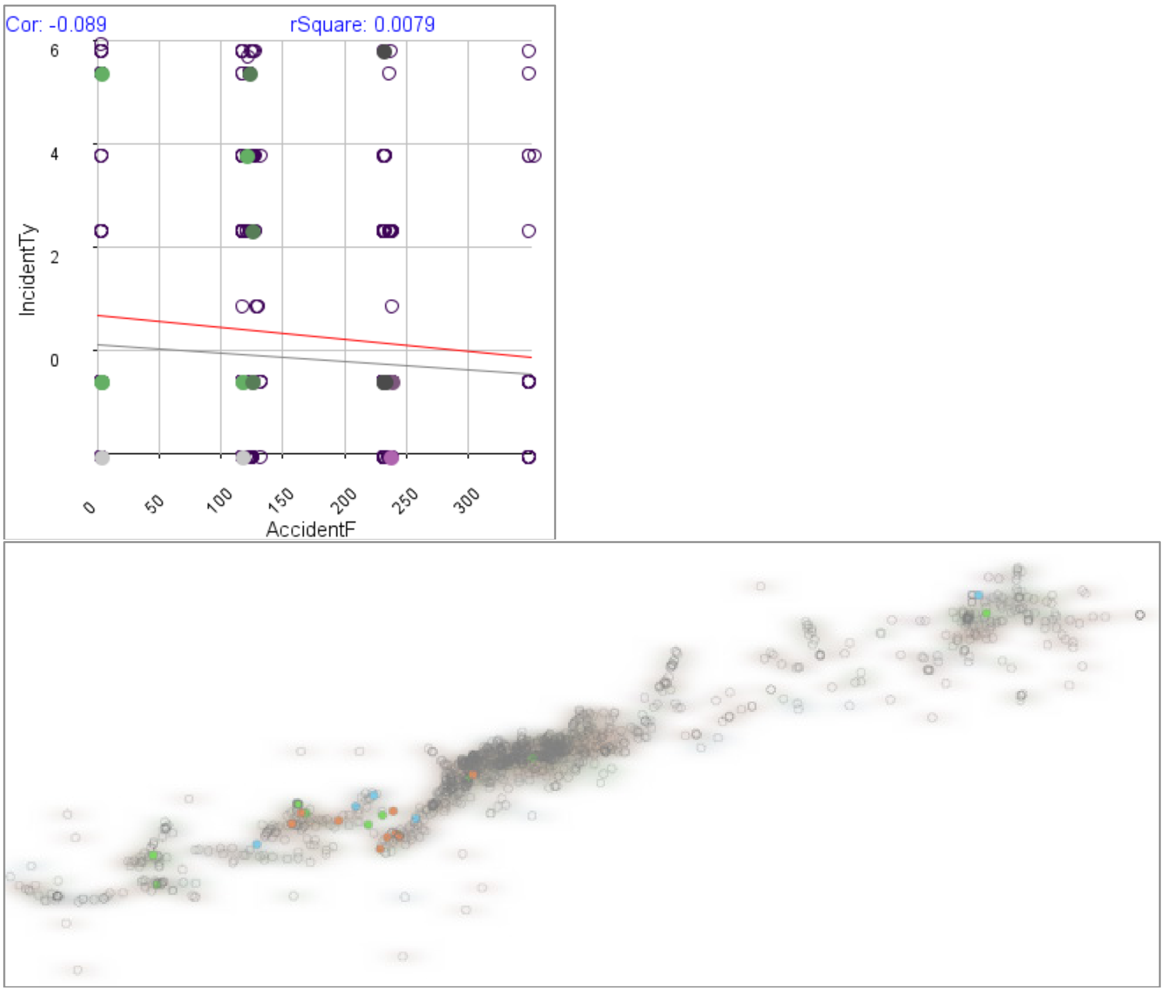


Figure 69 *Accident qualities during challenging weather conditions. Note that the values shown on scatterplot axes (upper window) do not match with the plotted items. See the value codes for IncidetnType in Appendix 2 and for AccidentFactor in Appendix 3.*

The scatterplot in the upper window (Figure 69) represents the incident types (vertical axis) of each accident factor (horizontal axis). Technical factors are generally in a need for assistance or rescue (code numbers for incidents see Appendix 2). One case is assigned as other incident type. The incidents in a case of accident/emergency are, besides *Assistance* and *Rescue*, also assigned as suspended and search operations as well as loss prevention. *Unknown* factors occur as *Assistance* or *Loss prevention* tasks.

The boat classes have not been investigated to this point in this analysis. New trends, such as jet-skies, may cause severe damages in high speeds. Many people have drowned in dinghies or in other smaller vessels. In Figure 70, the class that represents these types of recreational vessels are highlighted in the PCP (second leftmost axis) and plotted in the map window (lower window). The rightmost axis of PCP indicates the accident factors (value codes see Appedinx 2). There is a division between the incident types. The most common incidents are the blue polylines (*Rescue* and *Assistance* on the first leftmost axis) which indicate accidents caused by a technical factor. The other incident types are mainly suspended operations (pink). *Search*, and some other minor incident types (purple) are assigned as accident quality of *Accident/Emergency* (code numbers for accident qualities and accident factors see Appendix 3). It may be relevant to assume that such accidents happen on vessels, such as jet-skies, dinghies, kayaks, canoes, etc. Accidents on those vessel types may be fatal.

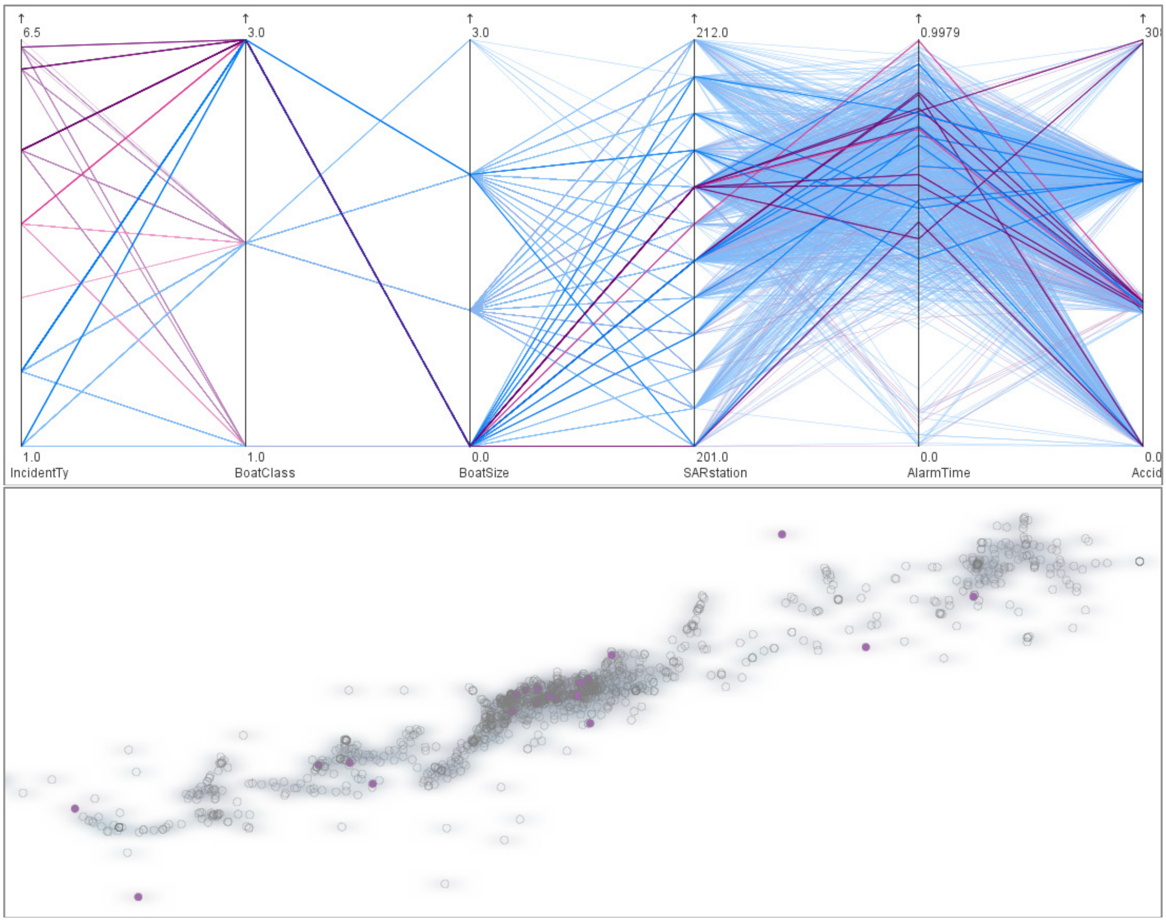


Figure 70 *Accidents on vessels of new trends (highlighted polylines) are either due to Technical or Accident/Emergency factors. This can be noticed from the rightmost axis. Value codes for Accident Qualities see Appendix 3. Attributes from the left in PCP (upper window): IncidentType, BoatClass, BoatSize, SARstation, AlarmTime and AccidentFactor. Value codes for the attributes see Appendix 2. Locations plotted in the bivariate map (lower window).*

The two major missions, *Assistance* and *Rescue*, do not have particular differences concerning accident factors and boat types. In the next step, the other incident types that have not been specifically investigated are mainly considered in the next examples. A mission type that was only addressed to the FLI is *Operation suspended*. This class is the third dominant mission type of this organization. The highest proportion value of *Operation suspended* occurred in the metropolitan area (see Appendix 5 (2/2). Figure 71 has the missions of *Operation suspended* highlighted in the PCP (upper window) and in the map window (lower window). In the PCP, the polylines are colored based on the type of boat: blue – sailboat, pink – motorboat, dark purple – other. All of these types of polylines are highlighted. The majority of the motorboats are the size of less than 7 meters, but also between 7 and 15 meters (third leftmost axis). The size for sailboats and other vessels is unknown. The missions did not occur during the night, but have occurred in the late evening before mid-night. The majority of the polylines of sailboats seem to intersect the factor group of *Accident/Emergency* whereas motorboats are involved mostly with technical factors. The polylines of other vessels types are vanishing in the clutters. The map shows that not many vessels of minor groups (dark purple) have been involved in suspended operations.

The colors on the map plot are based on the boat classes as well: light blue – sailboat, dark blue – motorboat, dark purple – other. Motorboats are located along the whole coast of the gulf. Cases of sailboats are mostly situated in the metropolitan region and some in Inkoo area. When zooming in the rightmost axis of PCP, the main accident factor of *Accident/Emergency* class is grounding, and it is mainly for sailboats. Breakdown is the main factor of technical failures, but in this case, all options for technical factors occur as well as factors of *Accident/Emergency* category, but mainly for motorboats.

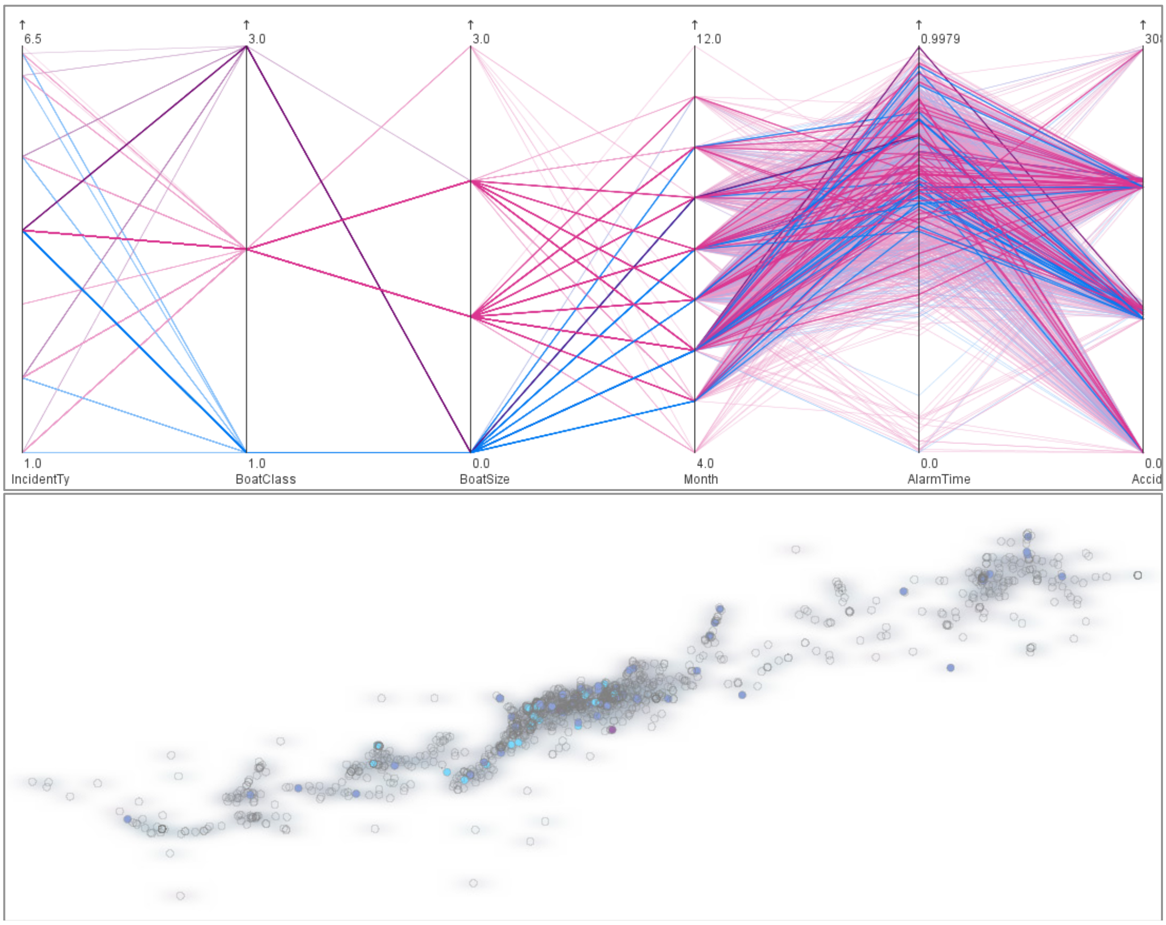


Figure 71 *Vessel types and accident factors involved in suspended operations. Attributes from the left in PCP (upper window): IncidentType, BoatClass, BoatSize, SARstation, AlarmTime and AccidentFactor. Value codes for the other attributes see Appendix 2, and for AccidentFactor see Appendix 3.*

In the next Figure 72, the pattern is investigated for *Search* group. The patterns is very similar to the case of *Operation suspended*, but more other vessel types are involved, located in Inkoo and metropolitan area (shown in the map as dark purple). When looking at the bivariate map (lower window), the situation of sailboats (light blue) seems to clutter in the metropolitan area, except one case is reaching the far West (Hanko). The cases of motorboats (darker blue) are cluttered in the metropolitan and Inkoo region. The PCP (upper window) shows the distribution between the sailboat and motorboats. Sailboats are clearly pointing at the *Accident/Emergency* class, whereas motorboats are pointing at all classes of accident quality. The sailboats again point at the *Accident/Emergency* group but this time, besides grounding, also due to MOB (man over board), tilt or fall, drifting vessel or object, getting lost or uncertain spot, and rescue of a human being. Cases of motorboat are pointing at the same factors in that group and at submersion or leak. The *Accident/Emergency* incidents on other vessels are pointing at MOB and 'getting lost or uncertain spot'. These single factors figured out by zooming in the axis *AccidentFactor* (rightmost axis).

***Figure** 72 Cases of search operations. Attributes from the left in PCP (upper window): IncidentType, BoatClass, BoatSize, SARstation, AlarmTime and AccidentFactor. Value codes for the other attributes see Appendix 2 and for accident factors see Appendix 3. Map window (lower): light blue: sailboat, darker blue: motorboat, purple: other.*

Technical failures (breakdown and propeller or power failure) again are common accident factors for motorboats. The uppermost accident quality class of the rightmost axis, *Other*, has highlighted polylines for missing person and loss prevention, mostly on other vessel types. Although, *Search* operations may be expected to provide aid and guidance on finding a missing person or helping a vessel lost in its way, the accident factor of a *Search* operation might be a factor or *Accident/Emergency*. However, it is not known how severe the incidents are.

Loss prevention, *Firefighting*, *Environmental accident/ prevention* and *Other* missions are highlighted in the Figure 73. The majority of the vessels are either sailboat (blue polylines in PCP and blue dots in a map) or other type of vessel (dark purple polylines in PCP and purple dots in a map). The sailboats are situated in Hanko, Tammisaari, Inkoo, metropolitan and Kotka regions. The other vessel types are located in the same areas except not in the East. Motorboats seem to be located only in metropolitan area and Tammisaari. The PCP (upper window) shows that the common factor for the most incidents of these incident types is unknown. The *Accident/Emergency* class has highlighted polylines on grounding, fire or explosion, submersion or leak, tilt or fall, drifting vessel or object, and getting lost or uncertain spot. Technical factors are not common for these incident types.

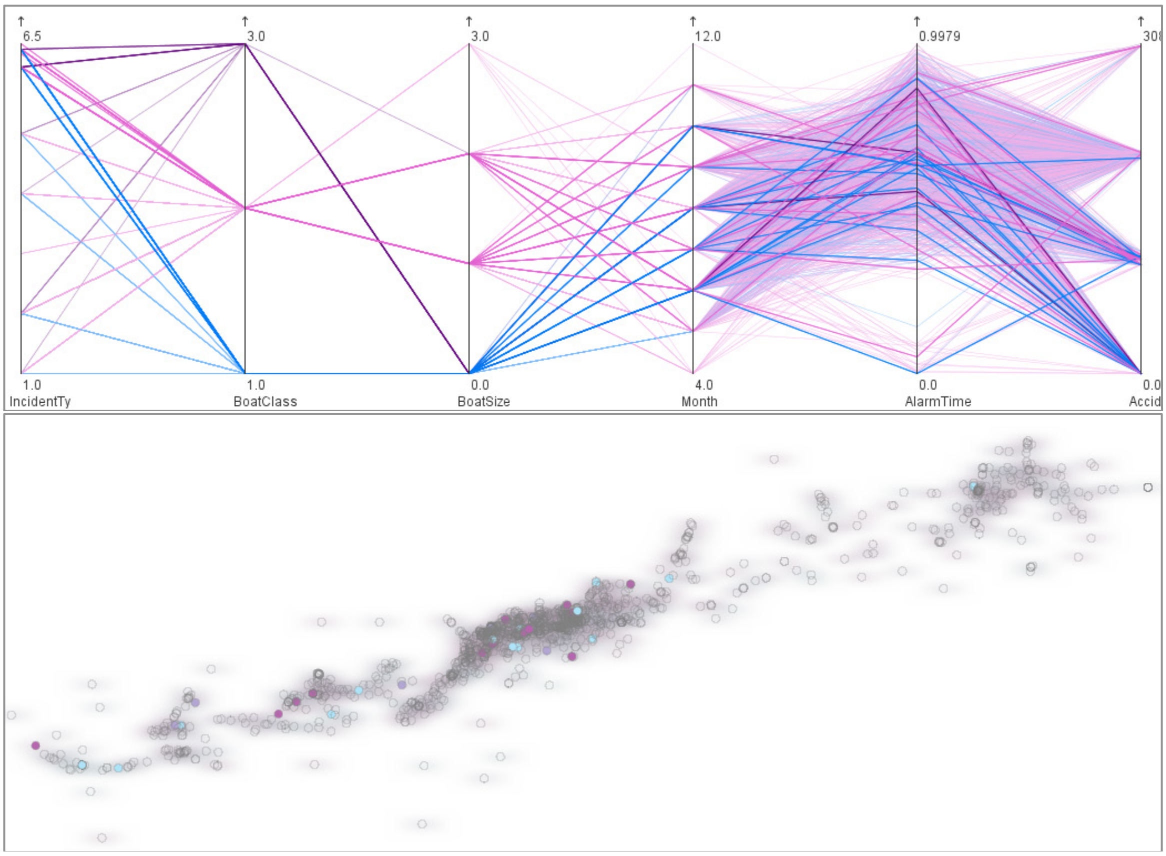


Figure 73 *Minor cases and their accident factors highlighted in the PCP (upper window). Attributes from the left: IncidentType, BoatClass, BoatSize, SARstation, AlarmTime and AccidentFactor. Value codes for the other attributes see Appendix 2, and for accident factors see Appendix 3. Map window (lower): light blue: sailboat, darker blue: motorboat, purple: other.*

After the investigation of accident factors for different boat types and incident types, we can state the different mission types may contain accidents due to several different factors. Some accident factors are more typical for certain types of vessels or accidents. Technical failures are the most common factors for motorboats, although accidents or injuries may occur. Sailboats might still continue traveling if a technical failure occurs with the aid of sails. Grounding seems to be the main factor of *Accident/Emergency* class and breakdown of *Technical* class. Accidents of other vessel types are mainly due to *Accident/Emergency* factors, which might indicate that the accident on those vessels might be fatal.

5.3 Examples of comparing SAR vessels

The FLI keeps statistics on the SAR vessels which perform the missions. The names and code numbers of the vessels are listed in Appendix 4 (1/2). In this section, the aim is to identify which SAR vessels are performing in challenging wave and wind conditions. In these examples, PCP technique is applied and the results are summarized to tables.

The most challenging weather condition was considered to have a high wind speed and wave height. The combination was selected efficiently by selecting the items from a single scatterplot. In Figure 74, those items are selected that reach approximately the risk limit of strong wind advisory (11 m/s) and one meter wave height. The PCP highlights the corresponding polylines. The SAR vessels can be identified on the rightmost axis of PCP (lower window). The identified SAR vessels, wind speeds, wave heights, and number of crew are listed in Table 6. The number of crew was checked from the websites of the FLI (Suomen Meripelastusseura 2014).

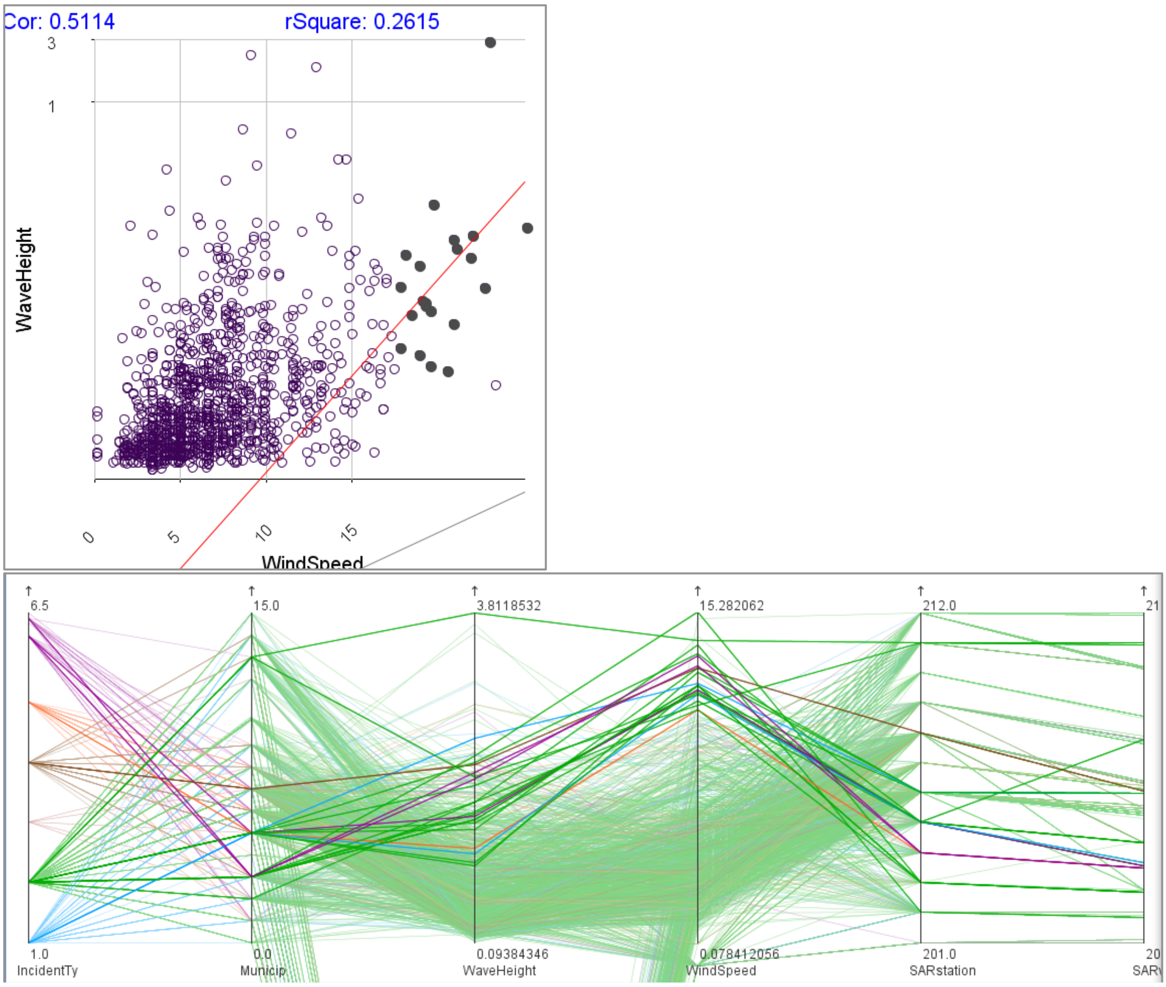


Figure 74 *Identifying SAR vessels that perform under challenging wave and wind conditions. Attributes from the left in PCP: IncidentType, Municipality, WaveHeight, WindSpeed, SARstaton, and SARvessel. Value codes for attributes see Appendix 2, for SAR vessels see Appendix 4 (1/2).*

Table 6 *SAR vessels that have performed in challenging wind and wave conditions.*

SAR Vessel	Home harbor	Wind speed (m/s)	Wave height (m)	Crew #
Ajax III	Tammisaari	**13,8**	1,67	3
Aktia	Porkkala	**13,4** \ 11,9 \ 11,2	2,12 \ 0,99 \ 1,44	4
Fagerö	Inkoo	**13,3** \ 10,8	1,93 \ 1,15	3
Kotka	Kotka	11,0	1,96	4
Lassi	Kotka	**14,0**	**3,80**	3
Mac Elliot	Porvoo	11,6	1,56	4
Mikrolog II	Espoo	12,7 \ 12,5 \ 12,0 \ 11,5	1,36 \ 0,95 \ 2,39 \ 1,85	2
PV225	Porkkala	11,5	1,09	2
Rautauoma	Helsinki	12,7	2,09	4
Reijo	Bågaskär	**15,3** \ 11,9 \ 10,8	2,19 \ 1,47 \ 1,68	2
Vega	Porkkala	11,7	1,54	?

Values that reach the limits of both heavy wind (14 m/s) and wave warning (2,5 m).
Average wind speed that is more than the strong wind advisory (11 – 13 m/s).

The locations and incident types, where the incidents happened, are already known from the previous sub-section 5.1.1 for all SAR organizations (Figure 55) and section 5.2 (Figure 69) for the FLI only. The column of home harbors of the vessels show an example which stations have been active. The PCP (Figure 74) shows that most of the SAR stations (axis *SARstation*) are located in the western area which also includes some other incident types besides *Assistance* (green polylines). The purple numbers in Table 6 show that only one vessel has managed to operate under a condition that reaches the minimum risk limit of heavy wind warning (14 m/s) and wave warning (2,5 m). This may indicate that the FBG is more active in severe circumstances. The green numbers indicate those wind speeds that exceeds the maximum risk limit of wind advisory (13 m/s). It may be concluded that the crew number of 3 and 4 may indicate bigger vessels, since they appear the most.

Although the combinations of the high wind speeds and wave heights can be easily selected with the scatterplot, the plot shows that there still remain items with high average wind speed and significant wave height, but still fall out of the selection criteria. For further exploration, those items have to be selected separately for each parameter. PCP technique was used in this part, which is illustrated in Figure 75. The coloring is according to the boat type (sailboat – blue, motorboat – brown, other – purple). Table 7 lists the vessel names and values of wind speeds and wave heights.

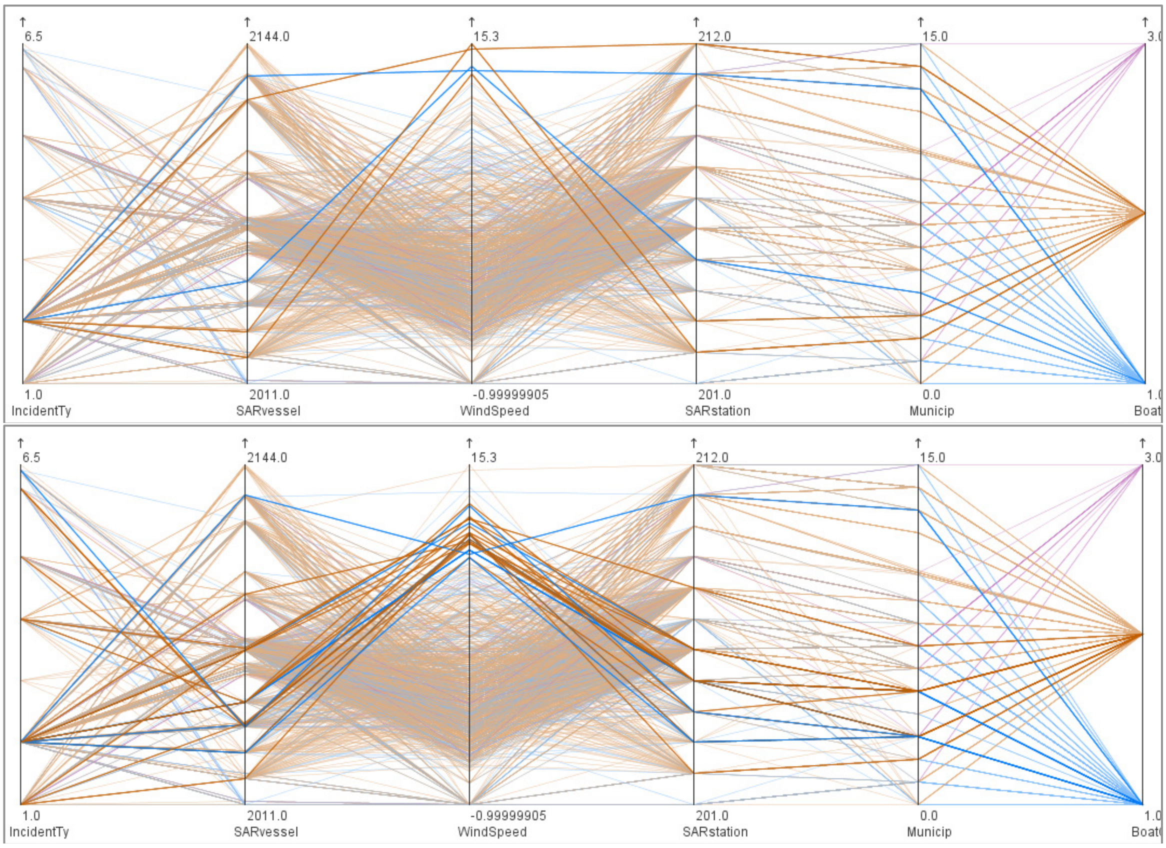


Figure 75 *SAR vessels that have operated in strong wind conditions: heavy wind (upper window), strong wind (lower window). Attributes from the left: IncidetType, SARvessel, WindSpeed, SARstation, Municipality, and BoatClass. Value codes for attributes provided in Appendix 2, for SAR vessels in Appendix 4 (1/2).*

Table 7 *Listed SAR vessels from Figure 75. Bolded numbers indicate values that reach approximately the risk limit of heavy wind (14 m/s).*

SAR Vessel	Home harbor	Wind speed (m/s)	Crew #
Ajax III	Tammisaari	**13,8** \ 12,0	3
Aktia	Porkkala	**14,2** \ 13,4 \ 12,3 \ 11,2	4
Fagerö	Inkoo	13,3 \ 12,8 \ 10,8	3
Hailikari	Hamina	**15,0**	3
Kotka	Kotka	11,0	4
Lassi	Kotka	**14,0**	3
Mac Elliot	Porvoo	11,6	4
Mikrolog II	Espoo	12,5 \ 12,0 \ 11,5	2
PV225	Porkkala	11,5	2
Rautauoma	Helsinki	12,7	4
Reijo	Bågaskär	**15,3** \ 11,9 \ 10,8	2
Savox	Tammisaari	11,9	?
Vega	Porkkala	11,7 \ 11,6	?

Table 7 lists the same SAR vessels as Table 6. The home harbor of the vessels again identifies where the conditions have been prevailing. Secondly, the PCP has blue and brown polylines highlighted, which indicate that sailboats and motorboats have been involved (the rightmost axis). The second rightmost axis reveals the municipality where the vessel or incident was located. Sailboats are identified to have been both in the West and in the East. It may be assumed that sailing requires wind and these conditions may have attracted sailors to set sail despite the warning.

The next example focuses on the wave warning. The selected polylines reach approximately the minimum risk limit of wave warning (2, 5 m) in Figure 76. Table 8 contains the results.

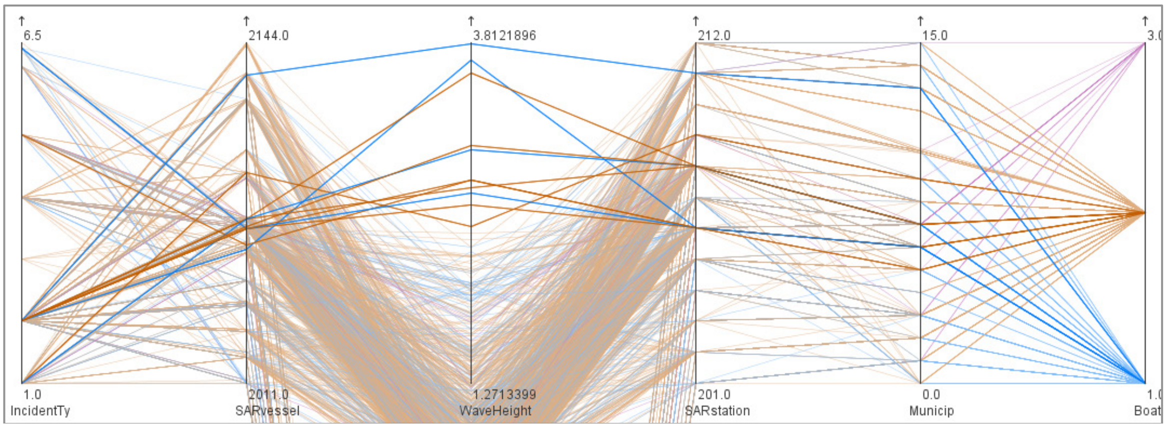


Figure 76 *Identifying SAR vessels operating under wave warning. Attributes from the left: IncidetType, SARvessel, WaveHeight, SARstation, Municipality, and BoatClass. Value codes for attributes provided in Appendix 2, for SAR vessels in Appendix 4 (1/2).*

Table 8 *Listed SAR vessels which have been on duty during the wave warning.*

SAR Vessel	Home harbor	Wave height (m) (*)	Crew #
Helvi Harjula	Kasnäs	2,78	2
Kuusakoski	Heinola	3,68	3
Lassi	Kotka	3,8	3
Mac Elliot	Porvoo	2,44	4
Mikrolog II	Espoo	2,78 \ 2,69 \ 2,6	2
Nihti	Hämeenlinna	3,58 \ 3,01 \ 2,73	2
Rautauoma	Helsinki	3,04	4

(*) Buoy of Helsinki

According to the results in Table 8, some of the home harbors are located in different municipalities that are not included in this study area: Kasnäs, Heinola, and Hämeenlinna. Those vessels may have been temporarily operating in the Gulf of Finland area. SAR vessel *Lassi* seems to be the only SAR vessel which has operated in the East (in Kotka). Other SAR vessels have operated in the metropolitan area, Espoo, Helsinki and Porvoo. Those SAR vessels with a crew number of 3 or 4 have been on duty when the wave height was high (in Helsinki buoy), but those with a crew number of 2 have also undertaken mission in high wave conditions. From this perspective, also vessels with a smaller crew number can operate in high waves.

5.4 Summary of key results and additional statistics

The aim of the exploration was to compare the incident occurrences and the SAR organizations under challenging weather and wave conditions as well as the radius of action and activity during different hours of the day. Possible relationships and interesting pattern such as outliers and exceptional values were under investigation. A more detailed analysis was applied to the data provided by the FLI, which included information on boat types, sizes and accident factors which were categorized into accident-quality classes. The aim was to find answers to the following questions:

- What are the general conditions when a certain type of an incident tends to happen?
- Under what conditions do the SAR organizations perform their missions? Are there any difference?
- Which visual analysis technique is suitable in discovering the patterns? Can we find any dependencies or outliers?

The finding of the first research question was that different incident types may happen in any environmental conditions. The first hypothesis was under investigation: "Wind speed and wave height are high during bad weather conditions". The dominant incident type during challenging wind and wave conditions was *Rescue* which was mainly obtained by the FBG. The second most common incident type *Assistance* was the most common for the FLI. These incident types appeared during high wind speeds and wave heights and during the combination of the two parameters. The FRS, who performed mainly *Ambulance and emergency responses* and some minor incident types (incident types 6.1−6.8, see Appendix 2), was active under these conditions as well, but not as often as the two other SAR organizations. The answer to the first research question is that the most common incident type of each SAR organization is most likely to happen: *Rescue* for the FBG, *Assistance* for the FLI, and *Ambulance and emergency response* for the FRS.

The coast of the Gulf of Finland was divided into three areas: the West, the metropolitan area and the East. The main locations of cases with both wind and wave warning were in the western and eastern parts of the Gulf of Finland. Especially Inkoo – Kirkkonummi (West) region had the densest clusters of incidents. Incidents also occurred in other regions but were not as densely situated as in the West. Metropolitan area had the most cases occurring when the wave warning was valid (~ 2, 5 m or higher). The heaviest wind speeds during operation have occurred in the West and somewhat in the East and rarely in metropolitan area.

The answer to the second research question was found when the performances of SAR organizations were compared. In the case of bad weather conditions, all three SAR organizations are active in any weather, but the FBG takes action the most. The FBG is the leading authority of maritime SAR activity and thus receives the highest number of incidents. The travel distances of the SAR organizations were compared. This part was focused on the hypotheses: "The radius of action is varying between the SAR organizations, meaning there are differences in travel distance between the missions and the SAR station or between the missions and the shore". The largest radius of action is taken by the FBG. The FBG is able to reach long-distance destinations that may be several kilometers away from the main land. The voluntary organization, the FLI, may also perform long-distance missions that might locate several kilometers away from the station but still is close to the shore. This could be identified because of the adequacy of the provided data on SAR stations. The maritime SAR cases obtained by the FRS were located normally nearby the shore, usually a couple of hundred meters away.

Table 9 shows some additional statistics of maximum and average wind speeds and wave heights in which the SAR organizations have been on duty. The maximum and average distances between incidents and SAR stations and distances to the shore are listed in Table 9. The percentage share of the doninant incident types for each SAR organization is included in the table as well. This table is included in this analysis to support the main findings.

***Table 9** Additional statistics.*

	Border Guard	Lifeboat Institution	Rescue Services
Max. wind speed (m/s)	15,8	15,3	16,3
Average wind speed (m/s)	4,7	4,3	4,8
Max. wave Height (m)	3,94	3,8	4,07
Average wave height (m)	0,76	0,72	0,68
Max. distance to shore (m \ NM)	24 576 \ 13	18 834 \ 10	2185 \ 1,2
Average distance to shore (m \ NM)	470 \ 0,3	285 \ 0,2	92 \ 0,05
Max. distance to closest station (km \ NM)	47 \ 25	**150 \ 81**	–
Average dist. to closest station (km \ NM)	12 \ 7	8 \ 4	–
Max. Distance to home station (km \ NM)	–	**230 \ 124**	–
Average dist. to home station (km \ NM)	–	12 \ 7	–
Major incident type	Rescue (91 %)	Assistance (83 %)	Ambulance and emergency response (32 %)

It was stated that the FBG is the most active organization during challenging wave and wind conditions. Table 9 shows that the other two SAR organizations are also on duty in such conditions. For example, the FRS has the highest maximum wind speed, average wind speed and the highest maximum wave height. The highest radius of action was identified for the FBG, which is shown by the maximum distance to the shore. However, the FLI has highest distances to the home station and to the closest station. The explanation may be that the FLI have conducted rehearsals or visits to other destinations. Home stations for the FBG incidents are unknown and FRS incidents are not provided with station data.

Proportions of the three major incident classes of the whole dataset were compared for the areas and for different hours of the day (Appendix 5 (1/2)). This part of the analysis was a spatio-temporal analysis of the dominant incident types. This part was focusing on the following hypotheses: "Most missions are performed during the day hours. What is the pattern during the night and late-evening hours?" *Rescue* class, which has the largest number of incidents, had a largest ratio in the East. The proportion value of *Assistance* class was noted to be highest in the metropolitan region, and *Ambulance and emergency response* in the West. The proportions of incidents during night and evening hours did not have significant differences. The incidents mainly take place during the day and early evening. During the day, the incidents may locate further away from the main land than during the evening and night. Evening hours were, in a visual perspective, more active of incident occurrence compared to the night hours, because people are rather boating still when it is bright than at night. For additional analysis, the proportions of the incident types have been calculated for the FLI (Appendix 5(2/2)).

It is obvious that SAR organizations stay active throughout the day, especially during summer months. It is known, that the activity of the FLI is reduced for the winder period. The boating season mainly concentrates on summer period. The proportions of incident types were compared in different parts of the Gulf of Finland, the West, the metropolitan area, and the East. The metropolitan area was identified to have the busiest and densest traffic.

The answer to the third research question is that clear dependencies between the incidents, SAR organizations and other attributes were not found, except between those which are trivial, for example, wind speed and wave height. The most applied technique in this analysis was PCP, because it can plot values of several attributes at once. For combination of the two parameters, scatterplot was selected from the Multiform Scatterplot Matrix. Scatterplot turns out to be suitable for selecting combinations, which is not possible with PCP or bivariate map. Bivariate map was a supportive tool for plotting the locations of the incidents. PCP and a scatterplot of Multiform Bivariate Matrix are suitable for plotting exceptional values and outliers as well as depicting correlations clearly. Bivariate map can mainly plot incorrect coordinates, in this analysis. PCP is easy and simple for the user. In addition, the variety of interaction is the widest in PCP.

The minor incident types were further investigated in a more detailed analysis for the FLI data. These incidents were chosen, because the amounts of incidents in *Assistance* or *Rescue* classes are large and do not give exceptional patterns. Tables 10 and 11 include supportive statistics for the more detailed analysis. The most common accident factor of *Accident/Emergency* class was grounding, which is common for *Rescue* incidents. Breakdown is identified to be the most common technical failure and typical for *Assistance* missions. Sailboats and other vessel types were mostly assigned as accident quality of *Accident/Emergency*. Note that Table 11 shows the proportion values of the vessel types for each Accident-Quality class not the amount. Accidents of motorboats were mostly due to technical factors. Sizes were only provided for motorboats. Size of less than 7 meters was the most common boat size.

Table 10 *Proportions (%) of FLI missions categorized in different Accident-Quality classes.*

	Unknown	Accident/Emeergency	Technical	Other	Major factor
Rescue	18	49	23	10	grounding (24%)
Assistance	16	23	60	1	breakdown (31 %)
Ambulance and emergency response	–	80	20	–	accident or injury (40%)
Operation suspended	10	42	46	1	breakdown (30 %)
Search	7	73	13	7	getting lost/ uncertain spot (27 %)
Other	40	47	8	5	unknown (40 %)

***Table 11** Percentage share of boat classes in different Accident-Quality classes.*

	Unknown	Accident/Emergency	Technical	Other
Motorboat	76	65	80	77
< 7 m	96	76	59	90
7−15 m	4	24	40	10
> 15 m	−	0,4	1	−
Sailboat	22	32	20	15
Other	1	3	1	8

Proportion values of boat sizes of motorboats in each Accident-Quality class.

6 Discussion

6.1 Discussion on SAR performance data

The aim of the work was to study characteristics and relationships between boating incidents and SAR organizations with secondary weather and wave related data. The data used in the analysis was provided by three different SAR organizations. In order to construct a shapefile of the relevant data, the data had to be preprocessed, which included data selection, filtering, reducing, transformation and integration. This part was a long-term phase, which is typical for large datasets of multiple sources. The enormous size of data files may have caused problems in updating the display or the response of the tool might not have been fast enough to response the user's interaction. This is decreasing the efficiency of the tool. Cluttering and overplotting of data items make data perception and tracking complicated. Therefore the possibility to interact with the data is necessary to get more insight and access to the data.

Secondary data related to weather and waves had to be acquired from the FMI. More data processing was required in order to extract the data values and link them to the incident data. Since weather and wave conditions are discrete and the weather observatories, where the weather data derive from, are not evenly located along the coast, so no accurate weather data could be linked to the incidents. In addition, the data of weather parameters had flaws depending on the observatory. The sampling time was varying between the weather observatories as well as amounts of gasps and missing values of the parameters. Secondly, the wave data was derived from one buoy only, which is located off Helsinki. Expert knowledge or advanced modeling techniques should have been applied to evaluate the data.

Since the weather and wave conditions are considered to have a significant impact on SAR operations, it would help a lot if the SAR organizations had statistics on the prevailing conditions. The FBG keeps track on the conditions, but it was not accurate or usable for this analysis.

The results show that the challenging weather conditions seem to have located particularly in the western Gulf of Finland. This can be referred to the fact that the western archipelago is attracting tourists from other areas. Secondly, people are spending more time on boats during summer. Destinations and harbors provide sufficient services, which makes boating more convenient. In the East, challenging weather and wave occurred as well. The eastern part of the gulf is a busy area for water traffic and international trade with narrow channels. The metropolitan area was identified to have the highest significant wave heights which derive from one buoy off Helsinki. According to the findings, equipment proposal to the western and eastern stations may be adequate response on both wind and wave conditions, especially wind because the western part is outermost to the open sea. For the metropolitan area suitable equipment would enable operations in challenging wave conditions. The significant wave heights derive only from one buoy. The significant wave heights of that buoy do not correspond the waves near the coast. This has to be taken into account. For example, Hanko is the most open area to the sea. Therefore, it is assumed that wind and wave conditions in Hanko may be more often challenging.

It was also identified, that the FBG may travel long distances to the destination, which might locate far away from the sore and from the closest station. The FLI may also travel long distances away from the station, but not that far away from shore. Depending on the case, the vessels need to reach the destination within a time limit, also for long distances. The maritime SAR missions are mainly obtained by the FBG and the FLI. The FRS may attend SAR missions in seas, but remain nearby the land and operates missions that are common on land. The distances were not provided with the incident data. Based on the coordinates and location information of the incidents and stations, the distances were calculated. For the calculations, cost surfaces had to be created.

Rescue, *Assistance* and *Ambulance and emergency responses* are the most common types of missions. *Rescue* missions are operated primarily by the FBG, *Assistance* by the FLI and cases in a need of ambulance or emergency response are obtained by the FRS. The activity of the SAR organization is high during the summer. They operate missions during other seasons as well, except the FLI may have a break during the winter season.

The available data of the FLI was consistent enough for a further analysis. The other reason to investigate the data of this organization more specifically was behind the atlas (Venäläinen & Sonninen 2013) that was made mainly for the missions of the FLI. The organization compiles information on vessel types and classifies the accident factors and accident qualities of the accidents. Size information is provided for the motorboats, which commonly are less than 7 meters. Boat sizes between 7–15 meters are quite common as well.

Accidents on smaller vessels may be fatal. New trends of vessels such as jet-skies can reach enormous speeds and thus cause severe damages. This applies for all engine-equipped vessels. Sailboat requires knowledge on controlling the vessel. It may be expected that boaters traveling with a sailboat could have more proper experience on boating than boaters traveling with motorboats. Motorboats are simpler to control than sailboats. Canoes, kayaks, rowboats, dinghies and boards may not require pre-knowledge on boating technique. Therefore, people on those vessels may lack adequate skills. Secondly, in Finland it is required to have adequate safety device for every person on board. However, it is not required to wear them.

According to the more detailed analysis, accidents on motorboats are mainly technical. Sailboats and other recreational vessels may have an emergency when an accident happens. This study, however, did not consider the human error, environmental disasters or human behavior on board.

Although the boating season in Finland is short, the bright evenings, national holidays and annual events will get people to have vacations and spend time on boats in a party purpose. It can be seen in the statistics that June and July are the busiest months of boating. It is highly recommended that people use safety devices, which are adequate for every person on board.

6.2 Discussion on techniques

In this analysis, the following visual analytic techniques were applied: parallel coordinate plot (PCP), bivariate map, and Multiform Bivariate Matrix. The usability of the techniques was compared to each other and their suitability for the task was evaluated.

The power of visual analysis tools is in the multi-linked dynamic views. Selecting items in one view highlights the corresponding items in another view. This enables the dynamic interaction. Table 12 lists the characteristics of the techniques that are implemented in GeoViz Toolkit, and which are rated according to a personal judgment from 1 to 5 (1 = poor, 5 = very good). The functionality, suitability and weaknesses of the implementation of the techniques in the toolkit are more specifically described in Appendix 6.

Table 12 *Rated properties of visual analytic techniques.*

	PCP	Bivariate map	MBM
Correlation	5	4	5
Trends	5	2	5
Outliers	5	4	5
Interactivity	5	3	3
Ease of use	3	5	4
Simplicity	4	5	4
Clarity of results	5	5	5
Ability to distinguish individual items	2	2	2
Multiple variables	5	1	5
Combinations of variables	1	2	5
Classification	4	4	1

Rates *1: poor, 2: passable, 3: adequate, 4: good, 5: very good*

The applied tools are all suitable for depicting correlations. Multiform Bivariate Matrix is powerful to plot correlations between every selected attributes. Thus it is easy to compare the trends of other pairs at the same time. However, it is not possible to change coloring themes, zoom in or out, or select a classification criterion with the Multiform Bivariate Matrix. Single scatterplots were picked form the matrix, because the bivariate maps in the lower window were not informative enough. Multiform Bivariate Matrix is able to plot multiple variables, but the larger the dataset is or the larger the number of attributes gets, the harder it is to interact with the tool. This applies to all these techniques.

PCP technique was applied in almost every part of the research. PCP was powerful by its interactivity options such as brushing, color coding, switching order of the axis, and zooming in and out. The user has to be familiar with the functionality of the tools before exploration. This technique is able to plot many attributes (six in this case) at once in a single view. The trend and correlations can be clearly and easily plotted but only for two adjacent variables (axes). It is not possible to select combinations of two variables like it is possible with the Multiform Bivariate Matrix technique. However, the values of other variables can be compared to each other. This technique was preferable in this analysis because of its simplicity and clarity.

Bivariate map works as an additional side tool for the other techniques. It is suitable for plotting locations of the selected items. Generally, the bivariate map technique is feasible for depicting the magnitude of correlation or proportion between two variables. The map is preferable for location-related data exploration rather than exploring multidimensional data. For example, outliers found with the bivariate map are mainly erroneous location values such as coordinates in this analysis. The technique does not clearly show the trend between two variables.

All techniques have problems with cluttered and overplotted items in a large dataset. This conducts to information loss. Many items easily get lost and are thus unrecognized. It is impossible to distinguish or select individual items among the data pile with these tools. Therefore working with multiple and linked views improves the data interpretation. However, the information loss cannot be fully avoided. The interaction tools have to be sufficient enough to order and classify the data. PCP and bivariate map techniques provided color coding and classification criteria. Multiform Bivariate Matrix did not have the option to select colors. The user was able to rearrange the value range for a single attribute in the Multiform Bivariate Matrix.

The advantage of the techniques is in their simplicity. They do not require specific expert knowledge. The results of the techniques are clear, which give an adequate general overview of the data. This data file was, however, too large to be explored in one piece with multiple views. Therefore the data file had to be split in order to work with the tools and explore the data deeper. How the tools are implemented in the toolkit is the main reason for the limitations. Usually, the problem is not in the technique itself.

It would ease the analysis and data processing in the future, if the SAR organizations agreed on a uniform data content and storage, and added easily observable meteorological data to the recorded data. This would make the data more comparable and easier to handle in data processing. In addition, it may not be needed to acquire much more additional data from other sources, which usually are in a different format.

7 Conclusions

The aim of this work was to investigate and compare maritime SAR missions and activity of SAR organizations during years 2007 – 2012 with additional weather and wave data by using the following techniques: parallel coordinate plot (PCP), bivariate map, and Multiform Bivariate Matrix. The analysis was a visual data mining exploration, where common visual analysis techniques were applied.

All techniques have been developed to depict correlations between two variables and to find outliers. In this dataset, the outliers were mainly erroneous or exceptional values or incorrect locations. PCP and a detailed scatterplot of the Multiform Bivariate Matrix were notably efficient tools for comparing relationships between the variables. PCP had sufficient interaction possibilities and a single scatterplot of the Multiform Bivariate Matrix was suitable for selecting combinations of two parameters as well as plotting the trend of correlation. Bivariate map was mainly used as a tool along with PCP or scatterplot. PCP was found to be the most comprehensive technique for this analysis.

When comparing the SAR missions and SAR organizations with the weather conditions, the main finding was that the incident types may occur in any weather at any time during the day related to general condition in the Gulf of Finland. The most investigated parameters were significant wave height and the average wind speed. The densest area of incidents under heavy wind and both wind and high waves was in the western Gulf of Finland. Metropolitan area was notably the busiest and densest area in water traffic in general with the highest significant wave heights of one buoy.

The SAR organizations stay active around the clock, but some differences in the performance between the SAR organizations were noticed. The FLI, as a voluntary SAR organization, reduces its activity for the winter season. The FBG and the FRS remain in a stand-by position throughout the year. Other differences were related to the traveled distances. The FBG has the widest radius of action, and thus the FBG is able to perform in the most challenging conditions in regard of the Gulf of Finland. The FRS undertakes their missions nearby land, which means that maritime SAR missions are mainly performed by the FBG and the FLI.

A more detailed analysis was applied using data provided by the FLI. The analysis investigated accident factors which were categorized into accident qualities. The data included information on boat types and boat sizes as well. The most dominant accident quality type was technical, usually due to breakdown or a power or input failure both on motorboats and sailboats in the most common mission types, assistance and rescue. Grounding and collision were the most common accident factors in a case of an emergency. Accidents of other types of vessels, such as canoes, kayaks, jet-skies, etc., were mainly identified to be caused by either a technical or an accident/emergency factor. An example for comparison of FLI SAR vessels was applied mainly using PCP technique.

References

Alenius, P., Myberg, K. & Nekrasov, A. (1998) *The Physical Oceanography of the Gulf of Finland: A Review*. Boreal Environment Research. Vol. 3. Pp. 97−125. ISSN 1239-6095.

Andrienko, G., Andrienko, N., Schumann, H., Tominski, C., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S., Jern, M. & Kraak, M.-J. (2010) *Space and Time*. In: Keim, D., Kohlhammer, J., Ellis, G. & Mansmann, F. (eds.) Mastering the Information Age: Solving Problems with Visual Analytics. Goslar, Germany: Eurographics Association. Pp. 57−86. [Referred 26.11.2013]. ISBN 978-3-905673-77-7. Available: http://www.vismaster.eu/book/.

Arola, T., Jalonen, R. & Kujala, P. (2007) *Meriliikenteen paikkatiedon tilastointi ja hyödyntäminen Suomenlahden meriturvallisuudessa*. Helsinki: Picaset. 69 pp. Kirjallisuuskatsaus, joka on tehty MS GOF-tutkimushankkeen esiselvityksenä. ISBN 978-951-22-8599-0. ISSN 1456-3045.

Axis Maps LLC (2009–2010) *Univariate Maps Versus Multivariate Maps*. [Online]. [Referred 21.1.2014] Available: http://indiemapper.com/app/learnmore.php?l=multivariate.

Azofra, M., Pérez-Labajos, C.A., Blanco, B. & Achútegui, J.J. (2007) *Optimum Placement of Sea Rescue Resources*. Safety Science. [E-journal]. Vol. 45:9. Pp. 941−951. [Referred 11.2.2014]. DOI: 10.1016/j.ssci.2006.09.002. ISSN 0925-7535.

Ball, T.J. & Eick, S.G. (1996) *Software Visualization in the Large*. IEEE Computer Society. [E-journal] Vol 29:4. Pp. 33−43. [Referred 22.11.2013]. DOI: 10.1109/2.488299. ISSN 00189162.

Bertin, J. (1967|1983) *Semiology of Graphics: Diagrams, Networks, Graphs*. Madison, Wisconsin, USA: University of Wisconsin Press. 415 pp. ISBN 978-58948-261-6.

Brown, A. & Feringa, W. 2003. *Colour Basics for GIS Users*. Harlow, England: Prentice Hall. 171 pp. ISBN 0-13-033343-3.

Carr, D.B., Littlefield, R.J., Nicholson, W.L. & Littlefield, J.S. (1987) *Scatterplot Techniques for Large N*. Journal of the American Statistical Association. [E-journal]. Vol.82:398. Pp. 424−436. [Referred 23.1.2014]. DOI: 10.2307/2289444. ISSN: 01621459.

Chen J., MacEachren, A.M., Guo, D. (2008) *Supporting the Process of Exploring and Interpreting Space-Time, Multivariate Patterns: The Visual Inquiry Toolkit*. Cartography and Geographic Information Science. [E-journal] Vol. 35:1. Pp. 33–50. [Referred 11.10.2013]. DOI: 10.1559/152304008783475689. ISSN 15230406.

Chen, J., MacEachren, A.M. (2008) *Resolution Control for Balancing Overview and Detail in Multivariate Spatial Analysis*. The Cartographic Journal. [E-journal]. Vol. 45:4. Pp. 261–273. [Referred 11.10.2013]. DOI: 10.1179/174327708X347764. ISSN 0008-7041.

City of Helsinki (2013) *Introduction to Helsinki City Rescue Department*. [Online]. [Referred 12.2.2014]. Available: http://www.hel.fi/hki/Pela/en/Introduction+to+Rescue+Department.

Deltamarin (2006) *Maritime Search and Rescue on the Eastern Gulf of Finland: A Study for City of Kotka*. Kotka: Deltamarin. 80 pp. Report for project 1032. [Referred 10.2.2014]. Available: http://www.merikotka.fi/julkaisut/REPORT_SARGoF.pdf.

Dix, A., Pohl, M. & Ellis, G. (2010) *Perception and Cognitive Aspects*. In: Keim, D., Kohlhammer, J., Ellis, G. & Mansmann, F. (eds.) Mastering the Information Age: Solving Problems with Visual Analytics. Goslar, Germany: Eurographics Association. Pp. 109−130. [Referred 26.11.2013]. ISBN 978-3-905673-77-7. Available: http://www.vismaster.eu/book/.

Elmer, M.E. (2012) *Symbol Considerations for Bivariate Thematic Maps*. In: Annual Meeting of North American Cartographic Information Society. Portland, Oregon, USA. 17–19.10.2012. NACIS. [Referred 6.1.2014]. Available: http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/278_proceeding.pdf.

Elmqvist, N., Dragicevic, P. & Fekete, J.-D. (2008) *Rolling the Dice: Multidimensional Visual Exploration Using Scatterplot Matrix Navigation*. IEEE Transactions on Visualization and Computer Graphics. [E-journal]. Vol. 14:6. Pp. 1141−1148. [Referred 23.1.2014]. DOI: 10.1109/TVCG.2008.153. ISSN 10772626.

Esri (1995–2012) *Creating the least cost path*. [Online]. ArcGis Help 10.1. [Referred 28.10.2013]. Available: http://resources.arcgis.com/en/help/main/10.1/index.html#/Creating_the_least_cost_path/009z00000021000000/.

Esri (1995–2011) *Cost Back link (Spatial Analyst)*. [Online]. ArcGis Help 10.1. [Referred 21.2.1014].Available: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z00000017000000.htm.

Esri (1995–2011) *How cost distance tools work*. [Online]. ArcGIS Resource Center. [28.10.2013]. Available: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_the_cost_distance_tools_work/009z00000025000000/.

Ferreira de Oliveira, M.C. & Levkowitz, H. (2003) *From Visual Data Exploration to Visual Data Mining: A Survey*. IEEE Transactions on Visualization and Computer Graphics. [E-journal]. Vol. 9:3. Pp. 378–394. [Referred 8.11.2012]. DOI: 10.1109/TVCG.2003.1207445. ISSN 1077-2626.

Finnish Geodetic Institute (2012) *Reference Systems for Coordinates and Heights*. [Online]. [20.2.2014] Available: http://www.fgi.fi/fgi/research/reference-systems-coordinates-and-heights.

Finnish Meteorological Institute (2014) *Warnings on Rough Waves*. [Online]. [Referred 31.1.2014]. Available: http://en.ilmatieteenlaitos.fi/warnings-on-rough-waves.

Finnish Meteorological Institute (2014) *Wind Warnings for Land and Sea*. [Online]. [Referred 31.1.2014]. Available: http://en.ilmatieteenlaitos.fi/wind-warnings.

Goerlandt, F., Torabihaghighi, K. & Kujala, P. (2013) *A Model for Evaluating Performance and Reliability of the Voluntary Maritime Rescue System in the Gulf of Finland*. In: Steenbergen, R.D.J.M, van Gelder, P.H.A.J.M., Miraglia, S. & Vrouwenvelder, A.C.W.M. (eds.). Safety, Reliability and Risk Analysis: Beyond the Horizon. London: CRC Press. Pp. 1351–1356. ISBN 978-1-138-00123-7 (printed) ISBN 978-1-315-81559-6 (electronic). (DOI: 10.1201/b15938-203).

Guo, D., Gahegan, M., MacEachren, A.M. & Zhou, B. (2005) *Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach*. Cartography and Geographic Information Science. [E-journal]. Vol. 32:2. Pp. 113−132. [Referred 11.10.2013]. DOI: 10.1559/1523040053722150. ISSN 1523-0406.

Han, J. & Kamber, M. (2006) *Data Mining: Concepts and Techniques*. 2nd ed. San Francisco, California, USA: Morgan Kaufmann. 800 pp. ISBN 13: 979-1-55860-901-3.

Hardisty, F. & Robinson, A.C. (2009) *The GeoViz Toolkit: Using Component-Oriented Coordination Method for Geographic Visualization and Analysis*. International Journal of Geographical Information Science. [E-journal]. Vol. 25:2. Pp. 191−210. [Referred 18.3.2014]. DOI: 10.1080/13658810903214203. ISSN 1362-3087.

Ilmatieteenlaitos (2013) *Aallot ja vedenkorkeus*. [Online]. [Referred 31.10.2013]. Available: http://ilmatieteenlaitos.fi/aallot-ja-vedenkorkeus.

Ilmatieteenlaitos (2013) *Merkitsevä aallonkorkeus*. [Online]. [Referred 28.10.2013]. Available: http://ilmatieteenlaitos.fi/aallonkorkeus.

Ilmatieteenlaitos (2013) *Salamat ja ukkonen*. [Online]. [Referred 29.10.2013]. Available: http://ilmatieteenlaitos.fi/salama-ja-ukkonen.

Inselberg, A. & Dimsdale, B. (1990) *Parallel Coordinates: a Tool for Visualizing Multidimensional Geometry*. In: Proceedings of the First IEEE Conference on Visualization (VIS'90). San Francisco, California, USA. 23−26.10.1990. Los Alamitos, California, USA: IEEE Computer Society Press. Pp. 361−378. ISBN 0-8186-2083-8. (DOI: 10.1109/VISUAL.1990.146402.)

Keim, D.A., Kohlhammer, J., Mansmann, F., May, T. & Wanner, F (2010) *Introduction*. In: Keim, D., Kohlhammer, J., Ellis, G. & Mansmann, F. (eds.) Mastering the Information Age: Solving Problems with Visual Analytics. Goslar, Germany: Eurographics Association. Pp. 1−6. [Referred 26.11.2013]. ISBN 978-3-905673-77-7. Available: http://www.vismaster.eu/book/.

Keim, D., Andrienko, G., Fekete, J.-D., Carsten, G., Kohlhammer, J. & Melançon, G. (2008) *Visual Analytics: Definition, Process, and Challenges*. In: Kerren, A., Stasko, J., Fekete, J.-D. & North, C. (eds.). *Information Visualization, Human-Centered Issues and Perspectives*. Berlin, Heidelberg: Springer. Vol. 4950. Pp. 154−175. ISBN 978-3-540-70955-8.

Keim, D., Panse, C., Schneidewind, J., Sips, M., Hao, M.C. & Dayal, U. (2003) *Pushing the limit in Visual Data Exploration: Techniques and Applications*. In: Günter, A., Kruse, R. & Neumann, B. (eds.). KI 2003: Advances in Artificial Intelligence, 26th Annual German Conference on Artificial Intelligence. Hamburg, Germany. 15–18.9.2003. Berlin, Heidelberg: Springer. Vol. 2821. Pp. 37−51 [Referred 11.11.2013]. ISBN 978-3-540-20059-8 (printed) ISBN 978-3-540-39451-8 (electronic). Available: http://bib.dbvis.de/uploadedFiles/141.pdf.

Keim, D.A. (2002) *Information Visualization and Visual Data Mining*. IEEE Transactions on Visualization and Computer Graphics. [E-journal]. Vol. 7:1. Pp. 100−107. [Referred 28.10.2013]. DOI: 10.1109/2945.981847. ISSN 10772626.

KMRC − Kotka Maritime Research Centre (2014) *RescOp*. [Online]. [Referred 18.3.2014]. Available: http://www.merikotka.fi/rescop/.

Kokko, A. (2013) *Sään aiheuttamat meripelastustehtävät Suomen meripelastusvastuualueella 2008−2012*. Master's Thesis. University of Helsinki, the Department of Physics. Helsinki. 93 pp.

Kraak, M.J., Ormeling. F. (2010) *Cartography: Visualization of geospatial data*. 3rd ed. Harlow: Pearson Education. 198 pp. ISBN 978-0-273-72279-3.

Kujala, P., Hänninen, M, Arola, T. & Ylitalo, J. (2009) *Analysis of the Marine Traffic Safety in the Gulf of Finland*. Reliability Engineering and System Safety. [E-journal]. Vol. 94:8. Pp. 1349−1357. [Referred 7.2.2014]. DOI: 10.1016/j.ress.2009.02.028. ISSN 0951-8320.

Leppänen, J.-M., Bruun, J.-E., Rantajärvi, E. & Salojärvi, J. (2012) *Suomen merenhoitosuunnitelman valmisteluun kuuluva: Meriympäristön nykytilan arvio*. Ministry of the Environment. 102 pp. [Referred 28.10.2013]. Available: http://www.ym.fi/download/noname/%7B62E78B95-3FCA-41ED-AC1C-DA865DF646AE%7D/34442.

Li, L. (2006) *Rescue Vessel Location Modelling*. MSs Thesis. Dalhousie University. Ottawa, Canada. 125 pp. ISBN 978-0-27592-4.

Liu, L. & Du, C. (2014) *Bivariate Mapping in ArcView GIS*. [Online]. [Referred 16.1.2014]. Available: http://proceedings.esri.com/library/userconf/proc99/proceed/papers/pap171/p171.htm.

Lunetta, P., Smith, G.S., Penttilä, A. & Sajantila, A. (2004) *Unintentional Drowning in Finland 1970−2000: A Population-Based Study*. International Epidemiological Association. [E-journal]. Vol. 33:5. Pp. 1053−1063. [Referred 11.2.2014]. DOI:10.1093/ije/dyh194. ISSN 1053-1063.

Luntiala, P., Komulainen, M.-L., Larsson, N.-E., Wallin, A. & Lyytikäinen, J. (2004) *Kuolemaan johtaneet veneilyonnettomuudet vuosina 2000–2002*. Ministry of the Interior/Police Department. 52 pp. Sisäasiainministeriön julkaisu 11/2004. ISBN 951-734-664-6 (printed) ISBN 951-734-665-4 (electronic).

MacEarchen, A.M. (2004) *How Maps Work: Representation, Visualization, and Design*. Paperback ed. New York, NY, USA: Guilford Press. 513 pp. ISBN 1-57230-040-X.

McKnight, A.J., Becker, W.W., Pettit, A.J. & McKnight, A.S. (2007) *Human Error in Recreational Boating*. Accident Analysis and Prevention. [E-journal]. Vol. 39:2. Pp. 398−405. [Referred 11.2.2014]. DOI: 10.1016/j.aap.2006.09.004. ISSN 0001-4575.

Mitina, N.N. (2006) *Properties of Oceans, Inland Seas, Coastal Zones, and Estuaries*. In: Kuhblaryan, M.G. (ed.). Types and Properties of Water. Oxford, United Kingdom: EOLSS. Pp. 281−310. ISBN 978-1-84826-922-4 (printed) ISBN 978-1-905839-22-3 (electronic).

Molberg, P. J., Hopkins, R.S., Paulson, J. & Gunn, R.A. (1993) *Fatal incident risk factors in recreational boating in Ohio*. Public Health Reports. Vol. 108:3. Pp. 340−345. ISSN 0033-3549.

Mäkilä, M. (2009) *Veneilyonnettomuudet ja niihin johtaneet syyt*. Thesis. Kymenlaakso University of Applied Sciences. Kotka. 32 pp.

NGA−National Geospatial-Intelligence Agency (2013). *World Geodetic System 1984*. [Online]. [Referred 28.10.2013]. Available: http://web.archive.org/web/20120402143802/https://www1.nga.mil/ProductsServices/GeodesyandGeophysics/WorldGeodeticSystem/Pages/default.aspx.

Norrington, L., Quigley, J., Russell, A. & Van der Meer, R. (2008) *Modelling the Reliability of Search and Rescue Operations with Bayesian Belief Networks*. Reliability Engineering and System Safety. [E-journal]. Vol. 93:7. Pp. 940–949. [Referred 11.2.2014]. DOI: 10.1016/j.ress.2007.03.006. ISSN 0951-8320.

OECD (2014) *Statistics eXplorer*. [Online]. [Referred 23.1.2014]. Available: http://stats.oecd.org/OECDregionalstatistics/#story=0.

O'Connor, P.J. & O'Connor, N. (2005) *Causes and Prevention of Boating Fatalities*. Accident Analysis and Prevention. [E-journal]. Vol. 37:4. Pp. 689–698. [Referred 11.2.2014]. DOI: 10.1016/j.aap.2005.03.011. ISSN 0001-4575.

Peterson, A.R. (2009) *Visual Data Mining: Using Parallel Coordinate Plots with K-Means Clustering and Color to Find Correlations in a Multidimensional Dataset*. Master's Thesis. Kutztown University of Pennsylvania, the Faculty of Computer Science. Ann Arbor, Michigan, USA. 52 pp.

Rajavartiolaitos (2014) *Finland's Maritime Search and Rescue System*. [Online]. [Referred 12.2.2014]. Available: http://www.raja.fi/sar/en/system.

Rajavartiolaitos (2014) *Maritime search and rescue (SAR).* [Online]. [Referred 12.2.2014]. Available: http://www.raja.fi/sar/en.

Rajavartiolaitos (2014) *Operations.* [Online]. [Referred 12.2.2014]. Available: http://www.raja.fi/sar/en/rescuers.

Rajavartiolaitos (2006) *Meripelastusopas: 1. Yleistä.* 1st ed. Helsinki: Edita Prima. 18 pp. ISBN 952-491-117-5.

Räsänen, J., Järvi, T., Mäkelä, K, Rytkönen, J., Hentinen, M & Hänninen, S. (2005) *Veneilyn määrä ja taloudelliset vaikutukset Suomessa.* Helsinki: Finnish Maritime Administration. 122 pp. Merenkulkulaitoksen julkaisuja 5/2005. ISBN 951-49-2101-1. ISSN 1456-7814.

Shneiderman, B. (1996) *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.* In: IEEE Computer Society (ed.). Proceedings of the 1996 IEEE Symposium on Visual Languages. Boulder, Colorado, USA. 3−6.9.1996. Los Alamitos, California, USA. IEEE Computer Society Press. Pp. 336−343. ISBN 0-8186-7508-X. ISSN 1049-2615. (DOI: 10.1109/VL.1996.545307).

Siirtola, H. & Räihä, K.-J. (2006) *Interacting with Parallel Coordinates.* Interacting with Computers. [E-journal]. Vol. 18:6. Pp. 127–1309. [Referred 28.12.2013]. DOI: 10.1016/j.intcom.2006.03.006. ISSN 0953-5438.

Sisäministeriö/Pelastusosasto (2014) *Pelastuslaitokset.* [Online]. [Referred 12.2.2014]. Available: http://www.pelastustoimi.fi/pelastustoimi/pelastuslaitokset.

Slocum, T.A., McMaster, R.B., Kessler, F.C. & Howard, H.H. (2009) *Thematic Cartography and Geovisualization*. 3rd ed. Upper Saddle River, New Jersey, USA: Prentice Hall. 561 pp. ISBN 978-0-13-229834-6.

Stikeleather, J. (2013) *The Three Elements of Successful Data Visualization*. Harvard Business School Publishing. [Online]. HBR Blog Network. 19.4.2013. [Referred 4.11.2013]. Available: http://blogs.hbr.org/2013/04/the-three-elements-of-successf/.

Stikeleather, J. (2013) *When Data Visualization Works – And When It Doesn't*. Harvard Business School Publishing. [Online]. HBR Blog Network. 27.3.2013. [Referred 4.11.2013]. Available: http://blogs.hbr.org/2013/03/when-data-visualization-works-and/.

Suomen Meripelastusseura (2014) *Suomen Meripelastusseura*. [Online]. [Referred 12.2.2014]. Available: http://www.meripelastus.fi/fi/suomen-meripelastusseura.

Thomas, J.J. & Cook, K.A. (2006) *Visualization Viewpoints: A Visual Analytics Agenda*. IEEE Computer Graphics and Applications. [E-journal]. Vol. 26:1. Pp. 10−13. [Referred 8.11.2013]. DOI: 10.1109/MCG.2006.5. ISSN 0272-1716.

Thomas, J.J. & Cook, K.A. (2005) *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, California, USA: IEEE Computer Society Press. 186 pp. ISBN 0-7695-2323-4.

Trafi (2014) *Veneilyturvallisuus*. [Online]. [Referred 12.2.2014]. Available: http://www.veneily.fi/veneilyturvallisuus.

Venäläinen, E. & Sonninen, M. (2013) *Suomen meri- ja järvipelastustehtävät – Karttoja ja analyyseja vuosien 2007–2012 tehtävistä*. Helsnki, Finland: Unigrafia. 439 pp. Aalto-yliopiston julkaisusarja TIEDE + TEKNOLOGIA 12/2013. ISBN 978-952-60-5326-4 (printed) ISBN 978-952-60-5327-1 (electronic).

Viertola, J. (2013) *Maritime Safety in the Gulf of Finland: Evaluation of the Regulatory System.* Turku: University of Turku. 64 pp. Publications of the Centre for Maritime Studies. ISBN 978-951-29-5490-2 (printed) ISBN 978-951-29-5491-9 (electronic).

Ward, M., Grinstein, G. & Keim, D. (2010). *Interactive Data Visualization: Foundations, Techniques, and Applications*. Natick, Massachusetts, USA: A K Peters. 513 pp. ISBN 978-56881-473-5.

Ware, C. (2013) *Information Visualization: Perception for Design*. 3rd ed. Waltham, Massachusetts, USA: Morgan Kaufmann. 536 pp. ISBN 978-0-12-381464-7.

Yao, X. & Jiang, B. (2005) *Visualization of Qualitative Locations in Geographic Information Systems*. Cartography and Geographic Information Science. [E-journal] Vol. 32: 4. Pp. 219 − 229. [Referred 17.1.2014]. DOI: 10.1559/152304005775194683. ISSN 1523-0406.

Zhang, J, Miller, L., Cook, D., Hardjasamudra, A. & Hofman, H. (2003) *Densityplot Matrix Display for Large Distributed Data*. In: Simoff, S.J., Noirhomme-Fraiture, M., Böhlen, M.H. & Ankerst, M.I. (eds.). Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003). Melbourne, Florida, USA. 19−22.11.2003. Sydney: University of Technology Sydney. Pp. 59−70. [Referred 21.4.2014]. Available: http://www.biostat.wisc.edu/~page/ICDM_Workshops/VDM-workshop.pdf.

Appendix 1: Codes for accidents used by the Finnish Lifeboat Institution

SMPS TAPAHTUMARAPORTTI

	HÄLYTTÄJÄ
011	Hädässä tai avuntarpeessa oleva
012	Ulkopuolinen henkilö
013	Hätäkeskus
014	Meripelastus-/lohkokeskus
015	Merivartiosto
016	Pelastuslaitos
017	Poliisi
018	Muu hälyttäjä
019	Oma havainto

	HÄLYTYSTAPA
021	Hätämerkki
022	Puhelin
023	Hälytysjärjestelmä
024	VHF puhe
025	VHF DSC
026	Virve
027	Muu

	SUORITE 1, 2 ja 3	
910	Pelastustehtävä	
911		*ihmisen pelastaminen*
912		*evakuointi*
913		*omaisuuden pelastaminen*
920	Etsintätehtävä	
921		*etsintä avomerellä*
923		*etsintä saaristossa tai järvellä*
924		*etsintä maalla tai saaressa*
930	Sairaankuljetustehtävä	
931		*sairaankuljetus alukselta*
932		*sairaankuljetus saaresta*
933		*ensivastetehtävä*
934		*sairaankuljetuksen avust.*
942	Sammutustehtävä	
943	Vahingontorjuntatehtävä	
944	Ympäristövahingontorjutatehtävä	
947	OSC-tehtävä	
948	Lentopelastuspalvelun avustaminen	
949	Tehtävä keskeytetty	
950	Muu tehtävä	
951		*tarkastus tai varmistustehtävä*
952		*vesiliikenneturvallisuustehtävä*
953		*eläintehtävä*
954		*huolto- tai kuljetustehtävä*
960	Avunantotehtävä	
961		*hinaus*
962		*karilta irrotus*
963		*apuvirran anto*
964		*polttoaineen vienti / täydennys*
965		*korjausapu*
966		*muu viranomaisen avustaminen*
967		*muu avunanto*
990	Kaupallinen tai ennaltaehkäisevä teht.	
991		*turvaamistehtävä*
992		*esittely-, neuvonta- ja koulutus*
993		*sidosryhmäyhteistyö, -harjoitus*
994		*kaupallinen työsuorite*

SMPS TAPAHTUMARAPORTTI

	ONN. TAI VAARATILANTEEN LAATU
100	Meri- tai vesiliikenneonnettomuus
101	*pohjakosketus tai karilleajo*
102	*yhteentörmäys*
103	*törmäys muuhun kuin alukseen*
104	*mies yli laidan (MOB)*
105	*tulipalo tai räjähdys aluksella*
106	*uppoaminen tai vuoto*
107	*kallistuma tai kaatuminen*
108	*ajelehtiva alus tai esine*
109	*eksyminen tai epävarma paikka*
110	*epäilty hätämerkki*
200	Tekninen vika
201	*konevika*
202	*potkuri- tai voimansiirtovika*
203	*purje- tai takilavaurio*
204	*peräsinvaurio*
205	*köysi tai vast. potkurissa /*
206	*polttoaineongelmat*
207	*muu tekninen vika*
300	Onn. tai vaaratilanne saaristossa tai
301	*rakennuspalo*
302	*liikennevälinepalo*
303	*maastopalo*
304	*muu tulipalo*
305	*vahingontorjunta*
306	*öljyvahinko / vaar.ain.aih*
307	*ihmisen pelastaminen*
308	*kadonnut henkilö*
309	*muut onnettomuustilanne*
400	Sairaankuljetus- tai ensivastetehtävä
401	*sairaskohtaus*
402	*tapaturma tai vamma*
403	*hukuksiin joutuminen*
404	*sukellusonnettomuus*
405	*raskaus*
510	Luonnononnettomuus
520	Ilmaliikenneonnettomuus
530	Tuntematon tilanne
540	Muu tilanne

	TAPAHTUMAN ENSISIJAINEN SYY
601	Tulipalo
602	Vuoto
603	Merimiestaidollinen virhe
604	Sairaus tai tapaturma
605	Yllättävä sään tai luonn.olosuhteiden
606	Virhearviointi tai huolimattomuus
607	Tekninen vika tai huollon laiminlyönti
608	Aluksen rakenteiden pettäminen
609	Alkoholi
610	Ilkivaltainen
611	Tuntematon
612	Muu syy
613	Ei avuntarvetta

	KOHDE
710	Matkustaja-alus
711	*yli 1000 henkilöä*
712	*200 - 1000 henkilöä*
713	*alle 200 henkilöä*
720	Lastialus
721	*tankkialus*
722	*roro-alus*
723	*kuivalastialus*
724	*muu lastialus*
730	Kalastusalus
731	*kalastusemäalus*
732	*kalastusalus 15 m tai enemmän*
733	*kalastusalus alle 15 m*
734	*muu kalastusalus*
740	Huvialus
741	*moottorivene alle 7 m*
742	*moottorivene 7-15 m*
743	*moottorivene yli 15 m*
744	*purjevene*
745	*soutuvene / -jolla*
746	*purjejolla / -lauta*
747	*kanootti / kajakki*
748	*vesiskootteri*
749	*muu huvialus*
750	Kulkuväline
751	*moottorikäyttöinen ajoneuvo*
752	*lautta tai vast. kellumaväline*
753	*muu kulkuväline*
754	*lentokone*
755	*helikopteri*
756	*muu ilma-alus*
760	Henkilö
770	Rakennus tai muu omaisuus
771	*asuinrakennus tai muu suuri*
772	*vapaa-ajanasunto*
773	*sauna tai muu pieni rakennus*
774	*maasto (alle 10x10m)*
775	*maasto (yli 10x10m)*
776	*maasto (yli 100x100m)*
777	*muu omaisuus*
780	Muu kohde

	JOHTOVASTUU
001	Meripelastuksen johtokeskus
002	Pelastustoiminnanjohtaja
003	Aluksen päällikkö
004	Poliisi
005	Lentopelastuskeskus
006	Ympäristöviranomainen
007	Muu viranomainen

Appendix 2: Table of classified attributes with nominal and numeric values

Category	Nominal	Numeric
Mission		
	Rescue	1
	Assistance	2
	Ambulance and emergency response	3
	Operation suspended	4
	Search	5
	Loss prevention	6.1
	Inspection/Back-up task	6.2
	Wildfire	6.31
	Building Fire	6.32
	Firefighting	6.33
	Other mission	6.4
	Environmental accident/prevention	6.5
	Animal rescue	6.6
	Executive assistance	6.7
	Traffic accident	6.8
Boat type		
	Recreational vessel	1
	Fishing vessel	2
	Other	3
Boat class		
	Sailboat	1
	Motorboat	2
	Other	3
Boat size		
	Size unknown	0
	Less than 7 m	1
	7–15 m	2
	More than 15 m	3
SAR organization		
	Finnish Rescue Services (FRS)	1
	Finnish Lifeboat Institution (FLI)	2
	Finnish Border Guard (FBG)	3
SAR station (FLI)		
	Hangon Meripelastajat	201
	Tammisaaren Meripelastajat	202
	Bågaskärin toimintakeskus	203
	InkoonMeripelastajat	204

	Porkkalan Meripelastajat	205
	Espoon Meripelastajat	206
	Meripelastusasema I	207
	Helsingin Meripelastusyhdistys	208
	Porvoon Meripelastajat	209
	Loviisanseudun Meripelastajat	210
	Kotkan Meripelastusyhdistys	211
	Haminan Seudun Meripelastusyhdistys	212
SAR station (FBG)		
	Hanko	301
	Tammisaari	302
	Porkkala	303
	Suomenlinna	304
	Glosholmen	305
	Kotka	306
	Haapasaari	307
	Hurppu	309
Municipality		
	External	0
	Hanko	1
	Raasepori/Tammisaari	2
	Inkoo	3
	Siuntio	4
	Kirkkonummi	5
	Espoo	6
	Helsinki	7
	Sipoo	8
	Porvoo	9
	Pernaja	10.1
	Loviisa	10.2
	Ruotsinpyhtää	10.3
	Pyhtää	12
	Kotka	13
	Hamina	14
	Virolahti	15
Weekend		
	Yes	1
	No	0
National holiday		
	Yes	1
	No	0

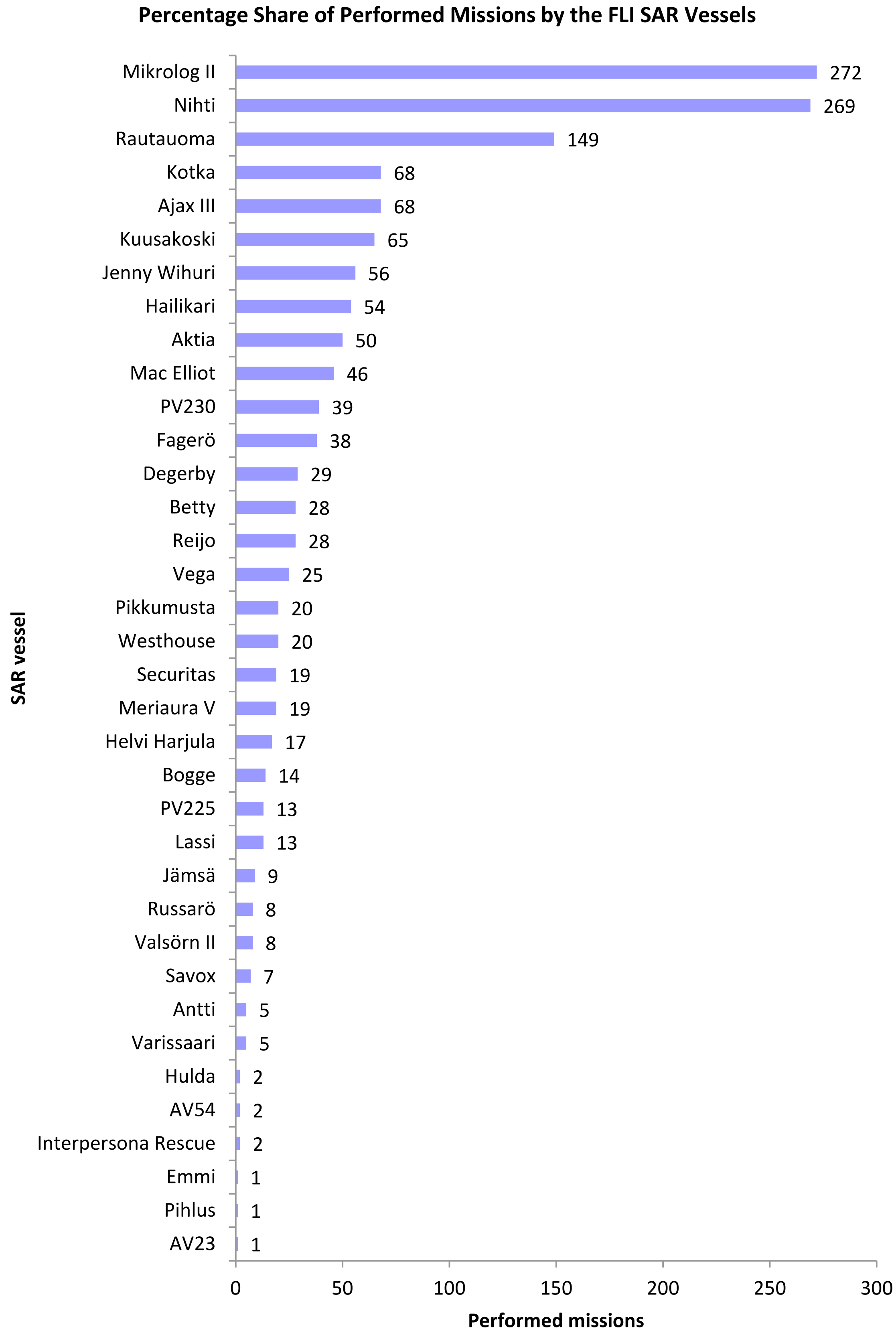
Appendix 3: Classified accident factors

100	**Accident/Emergency factor**	
101	Grounding	
102	Collision	
104	MOB (man over board)	
105	Fire or explosion	
106	Submersion or leak	
107	Tilt or fall	
108	Drifting vessel or object	
109	Getting lost or uncertain spot	
110	Suspected distress signal	
111	Sudden attack/Bout of illness	
112	Accident or injury	
113	Sinking on ice	
114	Rescue of human being	
115	First aid or evacuation	Total: 1 622
200	**Technical factor**	
201	Breakdown	
202	Propeller or power failure	
203	Sail or rigging damage	
204	Rudder damage	
205	Rope or similar object at propeller or rudder	
206	Running out of fuel or input failure	
207	other technical failure	Total: 2 727
300	**Other factor**	
305	Loss prevention	
308	Missing person	
309	Icing conditions	Total: 30
0	**Unknown factor**	Total: 1 032

Appendix 4: Table of SAR Vessels of the Finnish Lifeboat Institution, home stations, municipality of performance and the given code number, and mission performances

SAR vessel	Home Harbor	Municipality of Majority of Missions	Numeric
Ajax III	Tammisaari	Tammisaari	2021
Aktia	Porkkala	Kirkkonummi	2051
Antti	Helsinki	Helsinki	2074
AV23	Loviisa	Loviisa	2101
AV54	Porvoo	Porvoo	2092
Betty	Hanko	Hanko	2011
Bogge	Bågaskär	Sipoo	2032
Degerby	Loviisa	Loviisa	2102
Emmi	Espoo	Espoo	2061
Fagerö	Inkoo	Inkoo	2041
Hailikari	Hamina	Kotka/Hamina	2122
Helvi Harjula	Kasnäs	Espoo	2065
Hulda	?	Porvoo/Sipoo	2094
Interpersona Rescue	Tammisaari	Tammisaari	2022
Jenny Wihuri	Helsinki	Helsinki	2073
Jämsä	Jämsä	Hamina	2143
Kotka	Kotka	Kotka	2132
Kuusakoski	Heinola	Espoo	2063
Lassi	Kotka	Kotka	2131
Mac Elliot	Porvoo	Porvoo	2093
Meriaura V	Espoo	Espoo	2066
Mikrolog II	Espoo	Espoo	2071
Nihti	Hämeenlinna	Helsinki	2075
Pihlus	Rauma	Inkoo	2052
Pikkumusta	Hamina	Hamina	2121
PV225	Porkkala	Kirkkonummi	2043
PV230	Helsinki	Helsinki	2076
Rautauoma	Helsinki	Helsinki	2072
Reijo	Bågaskär	Inkoo	2031
Russarö	Hanko	Hanko	2012
Savox	Tammisaari	Raasepori	2023
Securitas	Porvoo	Porvoo	2091
Valsörn II	Espoo	Espoo	2064
Varissaari	Keuruu	Hamina	2144
Vega	Porkkala	Inkoo	2042
Westhouse	Espoo	Espoo	2062

Mission performance of the FLI SAR vessels



Appendix 5: Tables of proportions

Table 1 Proportions of mission for each area during different periods of the day. Numbers are in percents (%). The significant values are bolded.

	West	Metropolitan area	East
Proportion of the whole dataset	18,6	**63,9**	17,3
All in general			
Rescue	61,7	57,9	**69,7**
Assistance	23,6	**30,5**	23,6
Ambulance and emergency response	**6,6**	3,2	1,5
Operation suspended	0,8	1,8	0,9
Search	1,3	1,6	1,0
Other	5,8	5,0	3,4
Three major classes (Resc., Ass., Ambl.)	92,1	91,6	94,8
Night hours (0:00 a.m. − 5:00 a.m.)	4,3	3,2	4,2
Rescue	2,5	1,9	3,1
Assistance	0,8	0,5	0,6
Ambulance and emergency	0,7	0,4	0,1
Operation suspended	0,0	0,0	0,0
Search	0,0	0,2	0,1
Other	0,3	0,3	0,2
Three major classes (Resc., Assist., Ambl.)	4,0	2,7	3,8
Late-evening hours (9:30 p.m. − 0:00 a.m.)	6,4	6,9	6,3
Rescue	4,3	4,4	4,9
Assistance	0,6	1,7	1,2
Ambulance and emergency	4,7	0,2	0,0
Operation suspended	0,0	0,2	0,0
Search	0,3	0,2	0,2
Other	0,2	0,2	0,0
Three major classes (Resc., Assist., Ambul.)	**9,6**	6,4	6,1

Table 2 Proportions for the FLI missions for each area in percents (%). Significant ratios are bolded.

	West	Metropolitan area	East
Proportion of the whole FLI data	15,2	**71,4**	13,4
All in general			
Rescue	**10,3**	4,0	8,6
Assistance	76,2	83,9	**86,8**
Ambulance and emergency response	0,9	0,3	0,0
Operation suspended	3,6	**5,9**	4,1
Search	2,7	3,7	0,0
Other	**6,3**	2,2	0,5
Night hours (0:00 a.m. − 5:00 a.m.)	1,3	1,3	2,0
Rescue	0,4	0,0	0,0
Assistance	0,4	1,0	2,0
Ambulance and emergency	0,0	0,0	0,0
Operation suspended	0,0	0,0	0,0
Search	0,0	0,3	0,0
Other	0,4	0,1	0,0
Late-evening hours (9:30 p.m. − 0:00 a.m.)	4,0	5,8	4,6
Rescue	0,0	0,2	0,0
Assistance	3,6	4,7	4,6
Ambulance and emergency	0,0	0,0	0,0
Operation suspended	0,0	0,7	0,0
Search	0,4	0,2	0,0
Other	0,0	0,1	0,0

Appendix 6: GeoViz Toolkit

The GeoViz Toolkit is an environment for geovisualization. The user can build an environment by adding new views of analytical and exploratory tools interactively. It enables to explore the data through multiple views that are linked dynamically. Interaction in one view performs actions by the user and does highlighting in other views. In this study, the applied techniques are GeoMap, PCP and Multiform Bivariate Matrix. The user can pick color schemes and make classification settings for the data.

Unfortunately, the toolkit is quite old and some of the components are broken or unable to response to the user action. The tutorial does not provide broad guidelines for the tools. It only describes the components very briefly and not even all of the components that are provided. All the possibilities for the environment and settings the user has to figure it out by trying and playing with the tools. Another challenge for using the toolkit environment is in the efficiency in data exploration. The tools jam easily if the input shapefile is large. It works well if the amount of records is in hundreds. Thousands of records may cause jamming and response problems depending on the component.

The settings that the user has defined for the visualization criteria are only valid for the on-going session. For instance, if a certain color is given to a particular data class it remains only in that view for that current session. The toolkit is unable to save sessions, which means that the user cannot return to the previous working point of the analysis. The user has to be aware of the goal of the research and the functionality of the tools.

This chapter describes the applied techniques in the GeoViz Toolkit environment. The functionality and characteristics of interaction possibilities are explained with some snapshots. An article about the GeoViz Toolkit has been written by Hardisty & Robinson 2009.

Properties of GeoMap

The so called points or dots shown in the window are polygons with a ray of 1 km, which represent the point data as a graphical visual variable. The other significant visual variable is color. The map uses colors green and purple for the two selected variables (Figure 1). The gray scale with the color (saturation) indicates the correlation between the two variables: darker color indicates higher correlation, and lighter color weaker correlation. The classification, according to which the data is classified, can be selected in a dropdown box above the color bar. The list of classifiers includes the following options: *Quantiles*, *Modified Quantiles*, *Equal Intervals*, *Standard Deviation*, and *Raw Quantiles*. Quantiles method splits the number of items proportionally over the number of selected classes. This method is applicable when studying the correlations between different characteristics of the items, or when the areas are comparable by their size (Kraak & Ormeling 2010, p. 129). Modified Quantiles seems not differ from Quantiles. Equal Intervals is applied when the series of the items is linear each class occupying an equal interval along the number line. Standard Deviation shows how data is distributed along the number line. (Slocum et al. 2009. p. 69.)

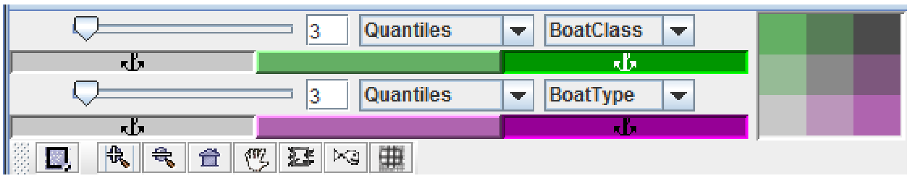


Figure 1 *Toolbar for GeoMap.*

The colors for the variables can be changed by double-clicking the color bar that is indicating a certain variable. Then a window pops up that provides colors of different styles (Figure 2).

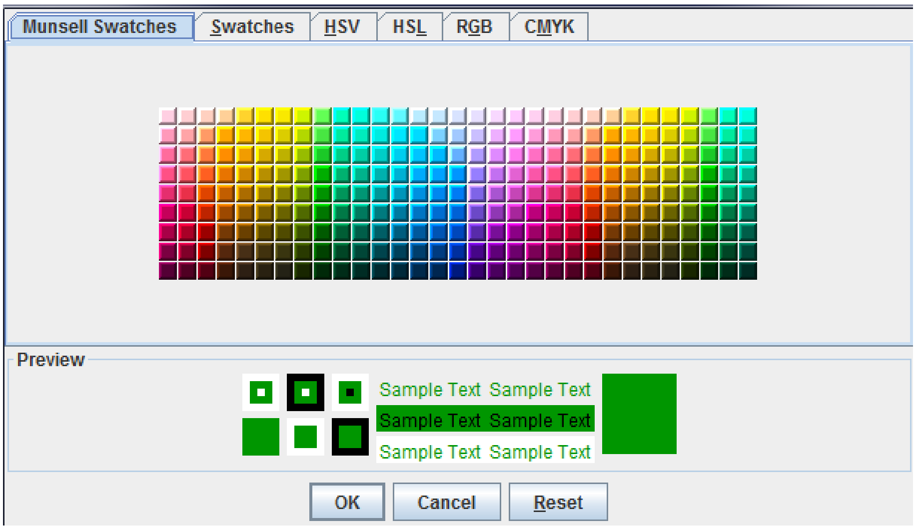


Figure 2 *Color palettes in the toolkit.*

The Gulf of Finland is West-East oriented, and the points of the missions are located along the coast line (Figure 3). To read the map, the surrounding has to be familiar. For example, only the points of the missions are shown, not the coast line of the main land. The metropolitan area, for instance, has to be known in advanced if the surroundings have to be explored with the map. Secondly, there are no borders of the foreign countries or international waters. As in the image it is shown that some outliers are located in the land area. This is hard to detect if the user is not familiar with the countries and the surroundings. The inland-located cases might possibly be in the river or in a lake if no additional information is provided.

With GeoMap it is possible to detect outliers related to location data. If some locations are clearly incorrect or interesting (also other than outliers), those can be selected by drawing a rectangular on the wanted dots, which remain clearly visible. Other dots, on the other hand, turn blurry and are not clearly visible anymore. A drawback for this step is that the selected dots cannot be compared to the other dots.

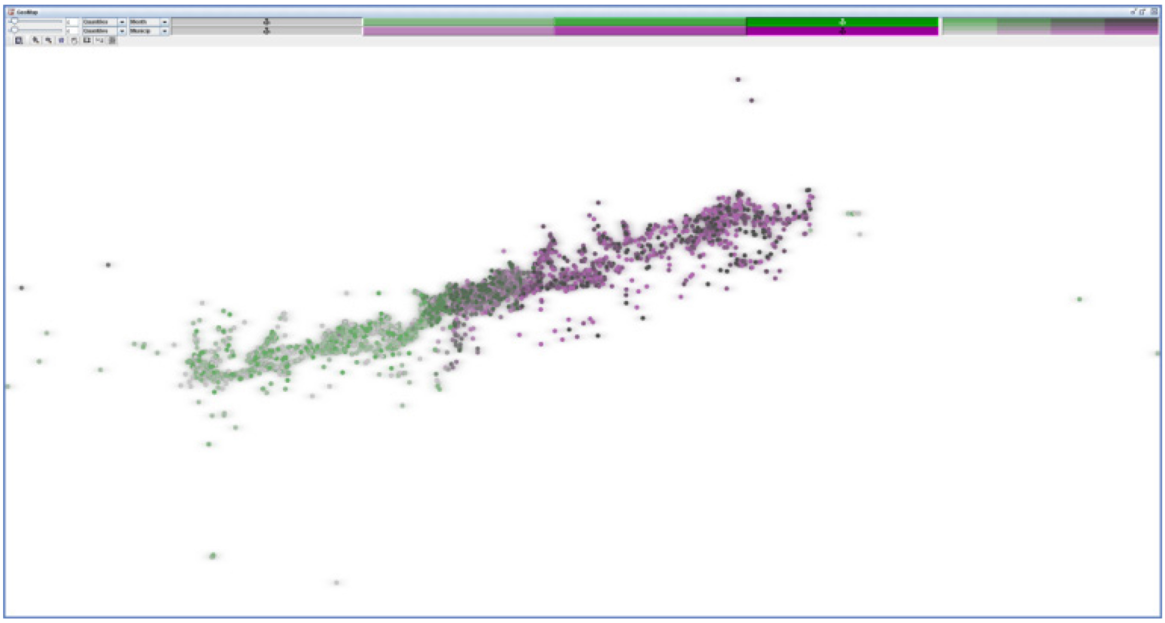


Figure 3 *The whole view of every incident how they are located along the coast.*

For receiving more information the toolbar provides eight different tools for data exploration (Figure 4). *Selection* mode, described previously, works by drawing a rectangular on wanted dots. The dots can be deselected by drawing a rectangular on an empty area on the background. *Zoom In* tool works similarly to the selection mode. It zooms in to the drawn rectangular, and also by clicking on the map. Zooming by drawing a rectangular is feasible in that sense that the selected area gets automatically centered in the view. With just a click the wanted view might relocate to the corner of near the edge, for example. *Full extend* zooms the whole map back to the view. *Pan* mode enables the user to drag the map in the window. *Excentric labels* is used to identify names of the neighbors. This tool is quite useless in this analysis, because the input data for this tool is completely numeric. This tool is useful for nominal data instead. Secondly, it is suitable for polygons that represent areas such as counties or municipalities. These dots are indicating incidents and thus have no specific names. *Fisheye lens* and *Magnifying Lens* have a similar function. They magnify the view within the circle they create on top of the map. The circle can be moved interactively by the mouse. They differ from each other in one thing. The scale within the Fisheye Lens is increased, but gets back to the normal size in the edges. The scale is also increased within the Magnifying Lens, but the scale remains the same, also in the edges. These two tools might be useful for exploring densely cluttered point data without zooming in. Also for exploring the borders of polygons more closely these tools may be appropriate. Once the Fisheye lens is selected, the mode can remain and it is difficult to deselect the mode. One to get rid of it is to pick the select mode or Excentric Labels.

Figure 4 *Tools provided for GeoMap (left to right): Selection mode, Zoom In mode, Zoom Out mode, Zoom to Full Extend, Pan mode, Excentric Labels, Fisheye Lens, and Magnifying Lens.*

With GeoMap it is feasible to detect cluttered areas. If the dots are very densely cluttered, the tools cannot give further information about the underlying dots (Figure 5). In order to get more details, more tools or techniques have to be provided.

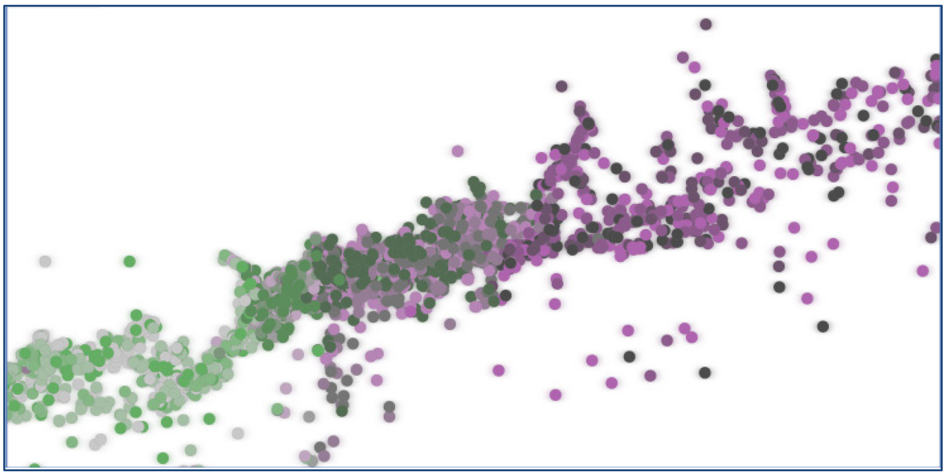


Figure 5 *Cluttered dots in the metropolitan area.*

The GeoMap in this this analysis is a bivariate map, which can only explore dependencies between two variables. For this reason, the pre-knowledge of the data is preferred when working with this dataset. Normally when dealing with bivariate data one variable is compared to the whole dataset, which the other variable represents. One example could be comparing the occurrence of an incident type in relation to the density of all incidents or population in a certain municipality. This kind of comparison is not available in this analysis.

Generally, it is challenging to distinguish the low saturated color values from one another in bivariate color sequences, which also occurred in this exploration. The background in the windows is white, which makes the light colors difficult to perceive. To solve this problem the background should be darker for stronger contrast. On the other hand, the darker the background gets, the harder it is to differ the high-saturated dark colors from the background. It is preferable, in this case, to change colors in the color bar but it might change the color outlook radically. The option to change the background color would be a supporting extension for the user to make an analysis. The legibility of the map is comprehensible if the number of classes is as low as possible, maximum of three or four. GeoMap is suitable for representing polygon data and compare areas. Data in this analysis is basically point data.

Properties of PCP

PCP plots polylines intersecting vertical axes. The axes indicate variables and one polyline is representing one data item. The point of the axis, where the polyline intersects, indicates the attribute value that is given to the data item. The user can select the variables from *VariablePicker* tool. The tool can be found from the menu bar under *Add Tool.* The tool is a window which lists the data attributes that are provided in the dataset. In this toolkit, the maximum number of variables is six which the PCP can plot at once. When the values are picked, *Send Selection* button, which is above the list, will be pressed and the PCP plots the polylines.

Under the plot, there is a list of tools (*Drag Mode*) that are used for data exploration (Figure 6). The tool options are *order*, *scale*, *translate* and *brush*. When *order* tool is selected, the user can change the order of the vertical axes by dragging one axis to the wanted position next to one or between two other axes. The correlation between the variables can only be seen between the juxtaposed axes. With the *scale* tool the user is able to zoom one axis bigger or smaller at a time. This works by pressing the left mouse button and dragging the cursor along the axes either up (zooming out) or down (zooming in). This method is feasible if polylines are intersecting the axes very closely to each other and when the intersections are hard to distinguish. It is also feasible to rescale the axes if some values or a value range is appropriate to leave out, for example, the values that represent null values.

After zooming in or out, it is preferable to adjust the values on the axes so that the value wanted value range would cover the axes. For example, if the main focus is a particular class and we want to exclude the values of other classes. *Translate* tool is applicable for this purpose. The user is then able to adjust the values in an appropriate position on the axes. If the polylines are densely cluttered, and it is not clear to recognize which polylines belong to which class, *brush* tool is the best solution for this problem. The tool highlights the selected polylines by dragging the cursor along the axis on that part that intersects the polylines. The unselected polylines get blurry and not that distinguishable anymore. This way the highlighted items can be compared to the whole dataset. Depending on the coloring of the polylines, the highlighted polylines will get a particular color. If the color for normal polylines is a dark color, for instance purple, the highlighted polylines turn blue. It also depends if the window is in a full-screen mode or not. If the highlight color is not blue, then they have that color what the user or software are defined to indicate them. The colors can be picked and changed in the same way as for the GeoMap. The visual classifier can also be determined in the toolbar (Figure 7).

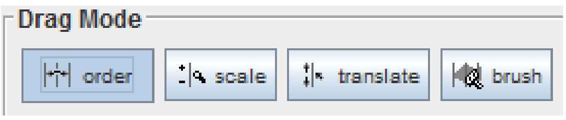


Figure 6 *Tool options for PCP.*

Figure *7 Toolbar for picking the visual classifier and color scheme for PCP.*

Another toolbar for data comparison is provided which includes options for scale adjustment: *0-max scale*, *min-max scale*, *min-max(abs) scale*, *and min-max variable* (Figure 8). The first option sets the values starting from zero. This applies for an analysis when only positive values are countable. A corresponding tool for negative values is not provided in the toolbar, which would be an appropriate extension, because for some null values in this dataset are indicated by negative numbers such as for the mission weather data. The second option, *min-max scale*, adjusts the whole value range visible for each axis. *Min-max(abs)* scale sets the values according to equal intervals. This way the value ranges can be compared to one another if the values have wide gasps and at which value range the most attribute data are pointing, for example (figure 9).

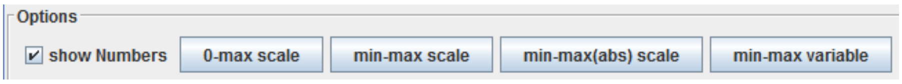


Figure 8 *Scale adjustment options.*

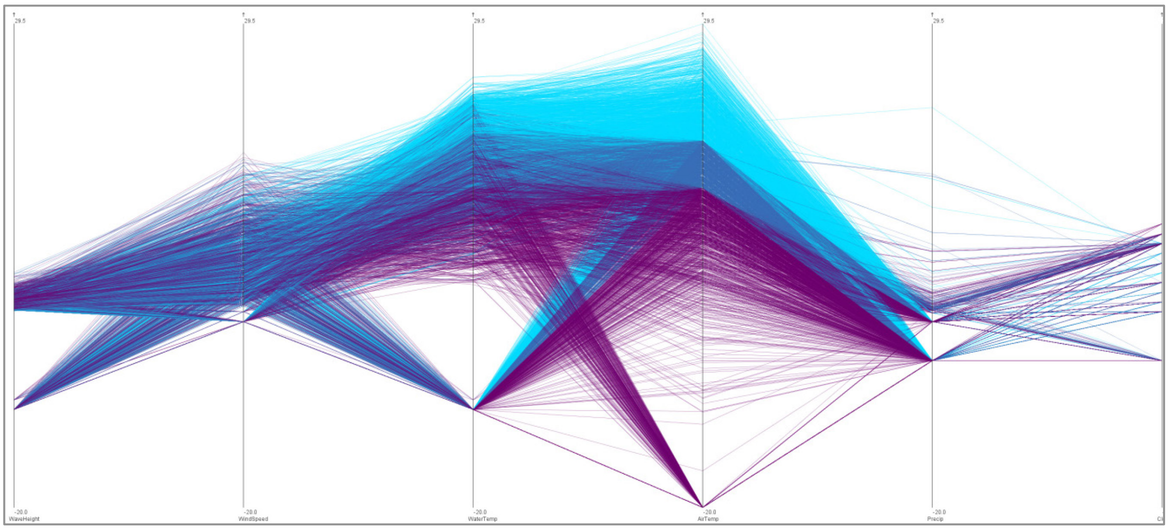


Figure 9 *min-max(abs) scale used to compare the weather and wave data. Wide gasps can be detected in the value ranges of attributes. WaveHeight (first left axis) and WaterTemp (3rd axis from the left) seem to have the widest gasps. Note that the lowest minus values indicate null values.*

The last scale adjustment option, min-max variable, retains the intervals of the values, and adjusts the values of the selected attribute (in the *Visual Classifier* toolbar) according to min-max scale. Only the value range of the currently selected attribute is shown as a whole. The other value ranges of other attributes are either partly or entirely shown in Figure 10. If the values exceed the values of the selected attribute, the values are excluded.

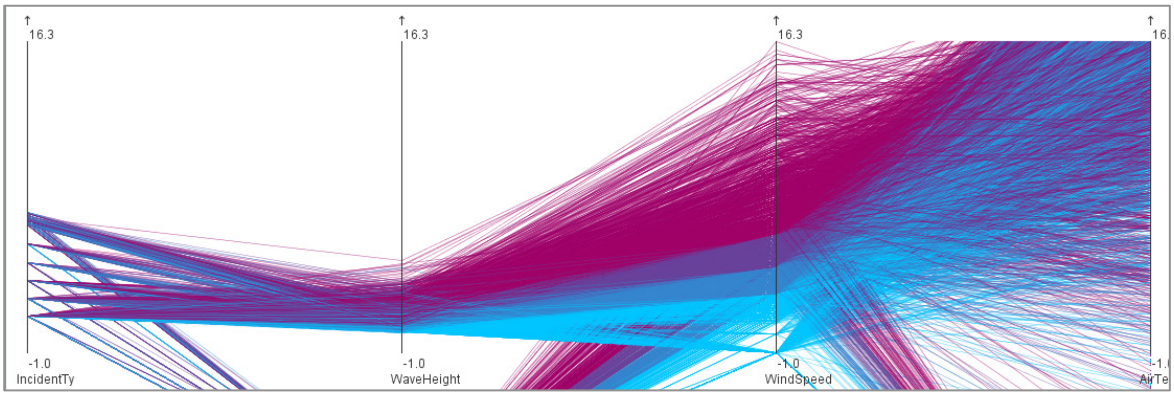


Figure 10 *min-max variable option used for the WindSpeed attribute (second rightmost axis). Values of other attributes (IncidentType and WaveHeight left, AirTemp right) that exceed the selected values are excluded.*

In general, the PCP works well and responses to the user's interaction. Depending on the size of the dataset and network connection, the usual flaw in the functionality, which may occur, is disappearing of the polylines when interacting with the data or rescaling the window. Once the brushing tool has been used and the background polylines are blurry, it is hard to deselect the tool and get the original outlook back. It works, for instance, by selecting a scale function or drawing a rectangle on another window, in a multi-view analysis, for example.

In a case when a class has subclasses, like the group containing the minor incident types, the values of each subclass might be very close to one another, which makes brushing difficult for a certain subclass. In this case, rescaling the axis is necessary. The use of the functionality for a first-time user is a bit confusing.

Outliers such as exceptional values, minimum and maximum, as well as gasps between the intervals are easy to detect with a PCP technique. With the cursor the user can point to a polyline and it will give the values of the points where the polyline intersect the axis. This helps if the brushing is complicated. The problem is that the cursor can point only those polylines that lie in the front. The polylines in the back are not visible even if the polylines are colored. The solution for this can be to make polylines of a particular class transparent, which can be found under the sub-window of the color scheme selection window presented in Figure 2 (the same window as for GeoMap), *HSV*. The selected coloring is valid during the session. The normal color hues go from black to purple. Overplotting and clutter are typical problems in a PCP representation. Therefore, appropriate tools are necessary for exploration.

Properties of Multiform Bivariate Matrix

The diagonal axis consists of histograms that show the amount of recorded items of each class of a certain attribute. The colors are the same both for the map and the points of scatterplot. The user, however, cannot change the color hue or the brightness. The given data is shown in scales of green and purple, one of the colors indicating one attribute. Also the grayscale is used for the level of correlation as in the GeoMap. The color of the bars of the histogram is gray and cannot be changed either.

In scatterplots a red regression line is drawn along the pattern in each grid. By right-clicking the background of a particular grid of scatterplot a menu bar will pop up. From the menu bar the user can select a function to modify the representation. The opportunities are resetting a range for the two variables (Set Range), edit, switching on and off the regression line (Regression), and switching on and off the selection line for selected items (Regression for Selection). The edit option does not function or do anything when selected. The user can also rearrange the order of the attributes by dragging the selected column or row to the wanted position. The other corresponding row or other column will automatically switch places. This is feasible when two attributes are being compared.

The items can be selected in the same way as in the GeoMap technique by drawing a rectangular on the wanted points. Besides the selection, the other functions work only for scatterplots, not to the bivariate maps. For histograms, none of the functions are valid.

The value ranges of *x*- and *y*-axes in scatterplots are not visible in the matrix. The user can double-click one scatterplot or a histogram, and the plot will appear in a new window including the values of the axes (Figure 11). The same interaction works similarly for the new scatterplot window as in the matrix.

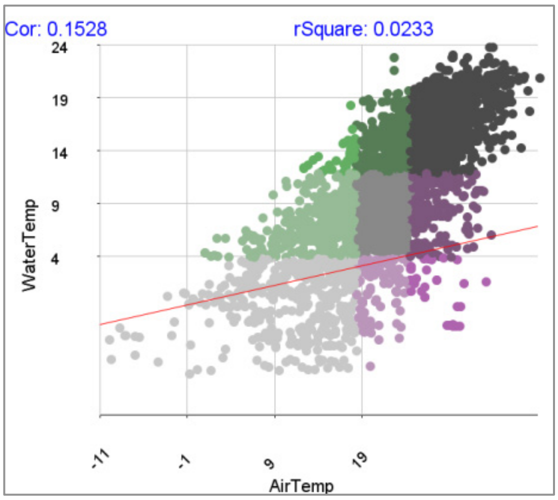


Figure 11 *Detailed scatterplot. There is a clear correlation between the air temperature and water surface temperature.*

The scatterplot for this dataset is suitable for finding clusters and outliers. It still suffers from overplotting of points which causes information loss. It describes the general behavior of the pattern and the relationship between two variables. Other advantage for this technique is that the user can select values that are combined by two different variables (figure 12). Other techniques applied in this study are not able to do so.

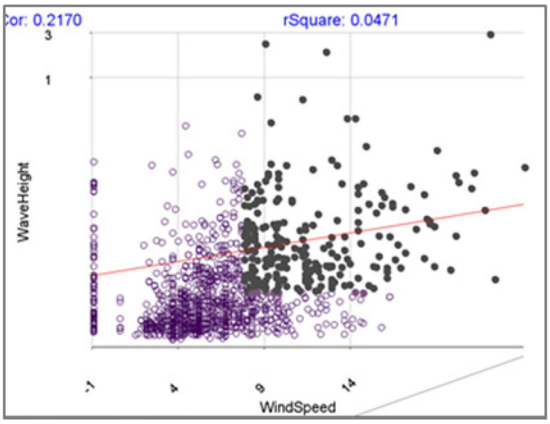


Figure 12 *Selected highest values of two variables (Wind Speed and Wave Height).*

For the Multiform Bivariare Matrix representation the user can pick the variables by clicking on the upper left corner of the matrix. However, the suitable number of variables for the matrix is the maximum of four, because the matrix will get jammed.

The patterns of the maps are the same as the bivariate map but in a smaller scale, which makes them hard to read (figure 13). Even zoom-in and -out functions are not provided for the detailed exploration of the maps. No window will appear by double-clicking either. One explanation is that the map represents point data of the incidents. A map representing the area consisting of polygons would be more appropriate for this technique. Other option is providing more advanced tools to improve the exploration.

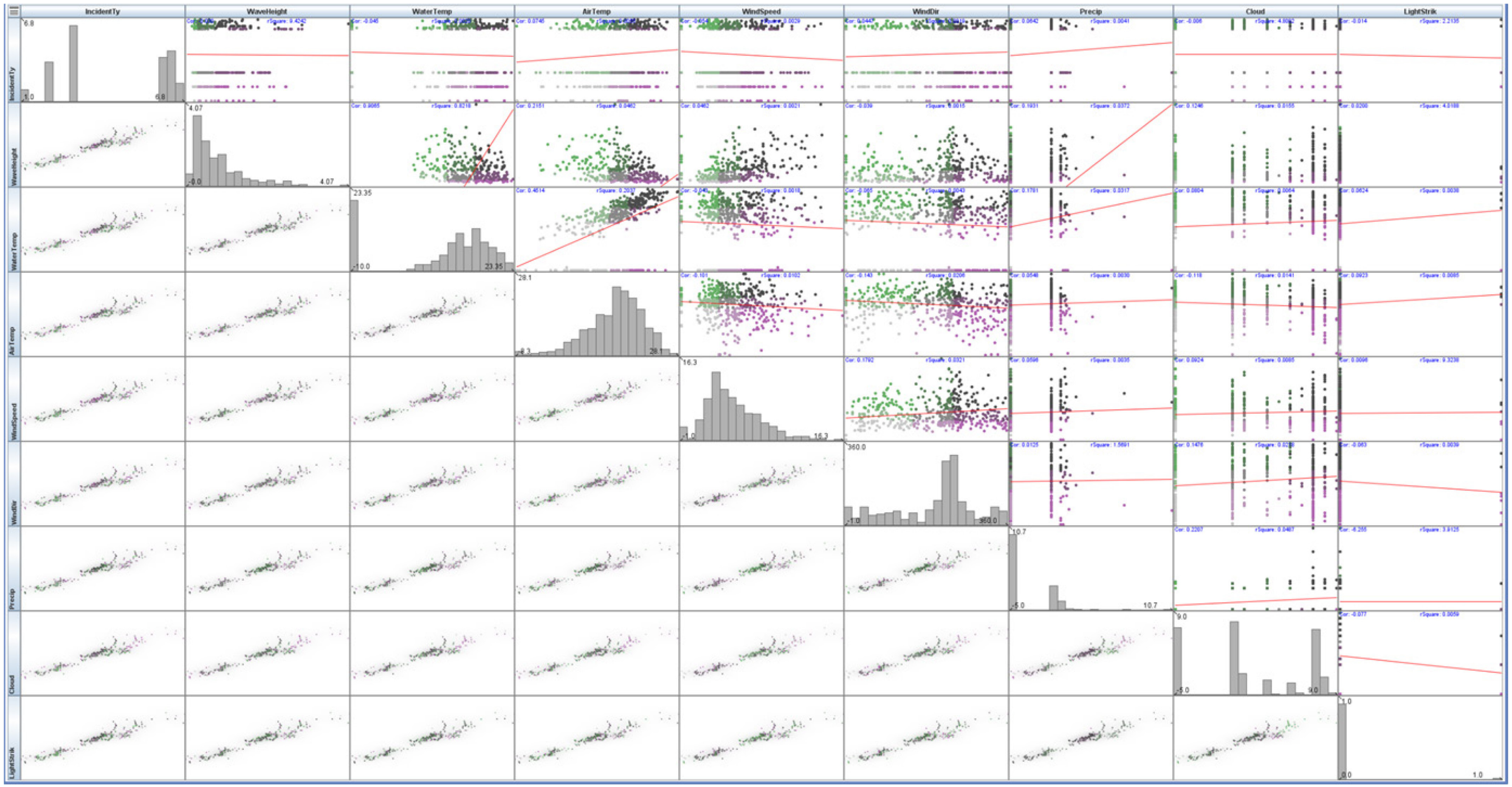


Figure 13 The map representations are too small for explore detail data. A Multiform Bivariate Matrix representation of FRS incidents and weather parameters.. No correlation seems to be between weather data and incident types (top row). Map windows in the lower part are very hard to read because of the small space. The lower part is not informative.