Nordic Journal of Surveying and Real Estate Research 9:1 (2012) 7–29
submitted on 28 November, 2011
revised on 14 May, 2012
accepted on 11 June, 2012

Exploratory vs. Model-Based Mobility Analysis

Jussi Nikander, Tanja Kantola, and Kirsi Virrantaus
Department of Real Estate, Planning, and Geoinformatics
School of Engineering
Aalto University
PL 11200, 00076 Aalto, Finland
Jussi.Nikander@aalto.fi, tmkantol@cc.hut.fi, Kirsi.Virrantaus@aalto.fi

Abstract. *In this paper we describe and analyze a visual analytic process based on interactive visualization methods, clustering, and various forms of user knowledge. We compare this analysis approach to an existing map overlay type model, which has been developed through a traditional modeling approach. In the traditional model the layers represent input data sets and each layer is weighted according to their importance for the result. The aim in map overlay is to identify the best fit areas for the purpose in question. The more generic view is that map overlay reveals the similarity of the areas. Thus an interactive process, which uses clustering, seems to be an alternative method that could be used when the analysis needs to be made rapidly and utilizing whatever data is available. Our method uses visual analytic approach and data mining, and utilizes the user knowledge whenever a decision must be made. The tests carried out show that our method gives acceptable results for the cross-country mobility problem, and fulfills the given requirements about the computational efficiency. The method fits especially to the situations in which available data is incomplete and of low quality and must be completed by the user knowledge. The transparency of the process makes the method suitable also in situations when results based on various user opinions and values must be made. The case in our research is from the crisis management application area in which the above mentioned conditions often take place.*

Keywords: *clustering, crisis management, explorative spatial data analysis, knowledge, multivariate visualization, spatial analysis process, visual analytics*

1 Introduction

1.1 Background and Motivation

The starting point for this research was the work done with the cross country mobility model and the uncertainty of the modeling results (Horttanainen & Virrantaus 2004, Virrantaus & Horttanainen 2004). The model in question has

been developed at the Engineering School of The Finnish Defense Forces and its purpose is to give an estimate about the mobility of the terrain in various climate conditions (summer, fall, winter, spring) for various vehicles. The model has been developed for many years (Orava 1997), and it has been verified by field tests and using expert knowledge. The model is now complete and both the source material and various parameters used in the modeling are available. For crisis management purposes, however, it has become necessary to modify the model in order to make it useful outside Finland, in different terrain, climate and data environments. It is natural that the existing model for Finnish terrain could not be used as such for example in Africa. A lot of model fitting work must be made. Furthermore, in areas for which data is not as readily available as in Finland, there are also numerous problems with data completeness and quality. The computational and human resources available in crisis management environments can also be limited. As a solution to this we have suggested a completely different approach, in which the traditional model is replaced by *an exploratory analysis process*, where computational and visual methods are complemented by user knowledge. Such a process is known as *visual analytic* approach (Andrienko et al. 2010, Keim et al. 2010, Thomas & Cook 2005). In our research, we have developed visual analytics methods that can be used as an alternative to the traditional model-based approach. We also compare our results to the existing model, and thus show that the new approach gives acceptable results and can be used for the analysis.

An extra motivation for this research was our experiences in a civilian crisis management exercise (MNE5, Multinational Experiment 5), in which our research group was involved in the development of a communication tool called SHIFT (Shared Information Framework and Technology) (Seppänen & Virrantaus n.d., Vesterinen 2008). In addition to collecting and sharing information there was a goal of utilizing analytic tools for various purposes. Some applications for the analysis were risk level estimation in the area, and cross country mobility analysis (Demšar et al. 2008, Zhang & Virrantaus 2010). In the experiment, it became very clear that in multi-actor and multi-agency activities, where actors do not always trust each other, the use of ready-made models to support decision making is not possible. Actors do not trust to the models developed by other agencies. Thus the need for a new type of methods that could show *transparency* and also *neutrality in the sense of being free from any pre-defined values,* was evident.

The focus of this work was in the cross-country mobility problem, but the results can be generalized to other types of analysis. The original mobility model, which was created by the Finnish Defense Forces, is based on a *map overlay* (O'Sullivan & Unwin 2003). The input data sets are layers of spatial data. Each layer is weighted according to how important it is for the problem, taking into account how the climate conditions changed throughout the year, and the data layers are combined to a mobility layer using map overlay. The model gives as the result the suitability classes for vehicle mobility. The development of the model was started in the 90's (Orava 1997) and it has been developed continuously. The quality of the input data has been analyzed and improvements to the model have been suggested (Horttanainen & Virrantaus 2004, Virrantaus

& Horttanainen 2004). The model is in operative use in Finland by the Defense Forces.

Cross-country mobility is an example of a more general problem of *suitability of an area* for a given activity. In this case areas of good mobility are suitable and areas of bad mobility are unsuitable. The map overlay indentifies the suitability of a given location by examining the input data values and taking into account the weights for the input layers, giving the result layer a suitability value based on this information. Thus, what we are actually interested in is identifying areas that *fulfill the given conditions.* Map overlay is just one possible strategy for identifying the suitability of locations. Another method is to use the concept of *similarity*. Since all suitable data items need to fulfill a given set of conditions, it is likely that these data items are similar to one other. For multi-variate data sets similarity can be calculated as a distance in multi-dimensional space. In case of non-correlating variables, simple Euclidean distance can be used. In the case of probable correlations distance measures like Mahalanobis distance can give more accurate results (Mahalanobis 1936). The suitability can then be solved by combining similar locations into classes, and giving each class a suitability value, since is self-evident that *similar areas are also equally suitable* to any purpose. This means that we can use a well known data mining method, *clustering,* to solve the problem (MacQueen 1967). In clustering the subareas are organized into classes according to their similarity. Various clustering algorithms exist. In our research we decided to use two simple and well-known algorithms: k-means (MacQueen 1967) and DBSCAN (Ester et al. 1996).

Clustering does not, however, directly solve the suitability problem. In a clustering result each class or *a cluster*, represents a set of similar data items. These clusters still need to be categorized according to the suitability of the items in the cluster. Thus, the clustering result needs to be interpreted. This can be accomplished using, for example, *multi-variate visualization* methods. In this work, we have decided to use the well-known *parallel coordinates plot* –method (Inselberg 1985).

As the constructive part of our research we have developed a prototype which has been tested in real data analysis situations. Our method results were compared with the traditional map overlay model by using expert evaluation and *misclassification matrix* (Zhang & Goodchild 2002) as measures.

1.2 Goals, Requirements and Limitations of the Research

The goal of this research is to develop a data mining and visual analytic approach that can replace the traditional map overlay based model for suitability analysis. The knowledge that is traditionally built in the model in the form of weights and parameters is now going to be inserted to the analysis process by the user. In our research we aim to develop a generic model of the suitability analysis process that is based on explorative and visual data analysis methods. We have developed our analysis model to be *free of inherent values or knowledge, transparent, user controlled, flexible, and simple to learn and use.* These requirements came originally from crisis management. We want to show that such analysis process

can be constructed by using well-known and simple computational and visual methods, and that the developed method can produce as good results as the traditional model-based approach.

The analysis process must be free of *inherent values or knowledge* in order to show no bias towards any actor in crisis management. Experience has shown that in crisis management, several actors refuse to use tools that can have bias towards other actors. The lack of inherent values or knowledge means that the user must be able to insert the values and knowledge required for a specific analysis during the process.

The analysis process needs to be *transparent*, which can have two meanings. First, users must be able see the details of how the analysis process is used, examine the available tools and ascertain that the process itself is free of inherent values and knowledge. Second, a specific analysis used to solve a given problem needs to be reviewable afterwards in order see the values and knowledge used, and see how the user has arrived from the input data to the analysis result.

The analysis process needs to be *general* so that an expert user can use it in various situations to solve a large number of problems. Decisions must be made on the data available and there is not much time to search for new or better data sets. Input data thus can be incomplete in many ways. The user must be able to compensate for the missing information by *his/her knowledge and risk taking*.

In this work we consider spatial problems that can be answered by dividing the given area into categories according to their suitability for a given activity. Unless otherwise specified, we divide the area into three categories: best-fit, suitable, and non-fit. Best-fit category covers the area which is best suited for the given activity, suitable covers areas that could be used for the activity, and non-fit areas that are unsuitable and cannot be used. The categorization is based on similarity, so we assume that when groups of similar areas are found they can then be ranked to the best fit, suitable and non-fit areas. The best categorization depends on the problem at hand, and the categorization described here was originally from our case example. A division into three categories is generic and thus useful when the goal of the analysis is to find places that are well-suited for a given activity. It can, however, be difficult to use such data in additional analysis. For such purposes, more categories can be useful, as more categories allow for more detail in the attribute dimensions.

Furthermore, in this work we're limiting ourselves to spatial problems where the input data can be transformed into a format where there are no explicit spatial dependencies between locations. Thus, the knowledge and information about spatial correlation between layers, as well as spatial autocorrelation between locations is not explicitly inserted into the process. If such knowledge is required, other computational methods or just user knowledge is used to analyze these phenomena. Spatial autocorrelation and the uncertainty of the analysis results of the customized model for terrain analysis have been analyzed in earlier research (Horttanainen & Virrantaus 2004). We limit ourselves to problems where the input data can be expressed in a raster format. Therefore, the suitability of a given location on a given input data layer for the given activity on can be expressed as

a simple attribute value. The reader should remember, however, that any vector data set can be rasterized.

The first of these limitations is due to the fact that we have observed it to be possible to solve a great number of spatial problems related to crisis management – as well as numerous other fields – by categorizing the area according to its suitability/similarity. The second and third limitations are placed in order to make it easier for us to construct a working prototype application. Our opinion is that the process we've described here can be used without these limitations, and investigating these claims is one topic for future research. The problems of data management and uncertainty issues have also been left out from this research.

1.3 Research Methods and Structure of the Paper

This paper can be divided into two parts: theory and implementation. In the first part a theoretical and conceptual approach is applied in order to investigate the nature of spatial data analysis process, and outline a *theoretical model* for the process. After this, in the second part, constructive approach is taken, and a prototype is designed and implemented in order to test our ideas in a real situation. The results of this experiment are reported and compared with the traditional model results.

The first part of the paper consists of sections 2–3. In Section 2 we describe related work in the fields relevant to this research. In Section 3 we introduce the conceptual model of the analysis process developed in this research. The second part of the paper consists of sections 4–5. Section 4 describes a detailed case example using cross-country mobility. Section 5 contains the experimental results gained from work on the cross-country mobility problem. It also contains a comparison of the results gained from our analysis process and the traditional model. Section 6 contains the discussion and conclusions.

2 Related Work

This research deals with several problems, main focus being in the role of knowledge input in the entire spatial analysis process, as well as the design of a values-free, transparent, effective and easy-to learn-and-use method for utilizing human knowledge.

Mathematical and computational methods that are used in spatial analysis process are typically documented carefully, but the human interaction and the "analytic discourse", the process between the analyst and the information (Thomas & Cook 2005), is often left without any attention. In the recent literature some leading researchers have pointed knowledge based phase of visual analysis process as one of the main topics in the visual analytics research. In the Research Agenda for Geovisual Analysis for Decision Support, published only few years ago (Andrienko et al. 2007), the major topics of visual analytics research are listed, and among them there is the following topic: *"Support of knowledge capture and manipulation"*. This means that the ideas that appear in the mind of the analysts should be put in form suitable for later review, communication to others and use in further analysis and in the subsequent phases of the analysis (Andrienko et al. 2007).

Another Research Agenda, published by The National Visualization and Analytics Center, USA is titled as "Illuminating the Path: The RD Plan on Visual Analytics" (Thomas & Cook 2005). It offers a framework and concepts for our research work under the title of analytic discourse. In the book McEachren and Kraak make recommendations that support our research plan: "1) Refine our understanding of reasoning artifacts and develop knowledge representations to capture, store and reuse the knowledge generated throughout the entire analytic process. 2) Develop visually based methods to support the entire analytic reasoning process, including the analysis of data as well as structured reasoning techniques such as the construction of arguments, convergent-divergent investigation, and evaluation of alternatives. These methods must support not only the analytical process itself but also the progress tracking and analytical review processes." Despite of these strong arguments towards the importance of dealing with the entire analysis process including knowledge input and management, it is not easy to find documented research on the topic. Most research work on spatial analysis and geovisualization are focused on selected methods and their development.

The visual representation of multidimensional or multivariate datasets in an understandable manner is a problem for which numerous different solutions have been proposed. Solutions include scatterplot matrices (Andrews 1972), star plots (Chambers et al. 1983), glyphs such as Chernoff faces (Chernoff 1973), reorderable matrices (Bertin 1981), and parallel coordinates plots (Inselberg 1985). When geographic data are being analyzed, such methods can be used to visualize the attribute data.

Parallel coordinates plot (PCP) (Inselberg 1985) is a multivariate visualization, which visualizes n-dimensional data using parallel axes in two dimensions. The axes are arranged either horizontally or vertically, and data points are visualized as line segments that traverse through these axes. PCP is limited to datasets where fewer than 1000 data items need to be shown on the screen simultaneously (Keim & Kriegel 1996). With larger data sets other visualization methods are required, or the amount of data visible needs to be restricted. Overall, however, PCP has been found to be a very useful tool which works well in combination with maps (Demšar 2006).

In order to analyze complex geographic problems visually, the user typically requires several different views of the data. One technique is to use *linked views*, where the user is given several visualizations and changes in one visualization are reflected in other views (Andrienko & Andrienko 2001, Chen et al. 2008, Edsall 2003, MacEachren et al. 1999). Such visualization systems have been proven to work in practice (Edsall 2003), and have been combined with computational analysis methods into successful tools (Chen et al. 2008, Guo et al. 2005, Sips et al. 2007). Such tools combine map visualizations, multidimensional data visualizations, and computational methods into one tool.

Clustering is the task of dividing a set of data items into a number of subsets, where elements in each subset are similar to each other, and elements in different subsets are distinct from each other. The similarity of the elements is calculated using a similarity measure. A simple Euclidean distance of data items is the

most common similarity measure. However, especially if the data have a lot of dimensions, other distance metrics, such as the Mahalanobis distance, which takes into account the correlation between input data dimensions, are also used.

K-means is a very well-known and relatively simple clustering method that can trace its origins to at least the 1960s (MacQueen 1967). K-means divides a set of data items into *k* clusters, where the number of clusters must be given beforehand. Each item belongs to the cluster with the nearest mean. K-means is often used in numerous different disciplines, and has a huge number of variations and improvements (Berkhin 2002). K-means has also been successfully applied in solving geographic problems. The algorithm has been used in, for example, finding good locations for facilities (Liao & Guo 2008), landslide hazard prediction (Gorsevski et al. 2005), and analyzing space-time paths (Shaw et al. 2008).

DBSCAN is a density-based clustering algorithm developed in the 1990s (Ester et al. 1996). The algorithm divides the data into clusters according to density: when sufficient number of data elements are close to each other, they form a cluster. Like k-means, DBSCAN is used in many different disciplines and has several variations. Perhaps the most important of the variations is GDBSCAN, which generalizes the concepts of density and neighborhood the algorithm uses (Sander et al. 1998). DBSCAN and its variations have been used to study, for example, clusters in road networks (Stefanakis 2007) and earthquakes (Pei et al. 2010).

There are also many other clustering methods which have been used in geoinformatics. For example, Guo uses hierarchical clustering methods in (Guo 2008), while Shoshany and others compare k-means to different clustering methods and other alternative strategies (Shoshany et al. 2007). Another common approach seems to be the use of self-organizing maps, such as in (Chen et al. 2008, Jiang & Harrie 2004).

3 The Model of the Analysis Process

In this section we introduce a model of an analysis process that fulfills the requirements we set in section 1. The problem we're trying to solve is the suitability of an area for a given activity. The result of the analysis will be classification of the area into categories, such as best-fit, suitable, or non-fit. Here we're concentrating on solving problems where there are no explicit spatial dependencies between input layers.

An overview of the analysis process can be seen in Figure 1. The left-hand side in the figure shows the model for the analysis process, and the right-hand side of the figure shows how input data is modified through the process. On the left-hand side of the figure is the analysis process, which consists of four main phases divided into a number of subphases. The boxes in the figure stand for different phases of the process, with the arrows between them indicating how the user can move from one phase of the process to another. The process flow is from top to bottom, and the backwards arrows from various phases of the process indicate how it is possible to work iteratively by moving back to a previous phase of the process. There are no backwards arrows from the computational data analysis

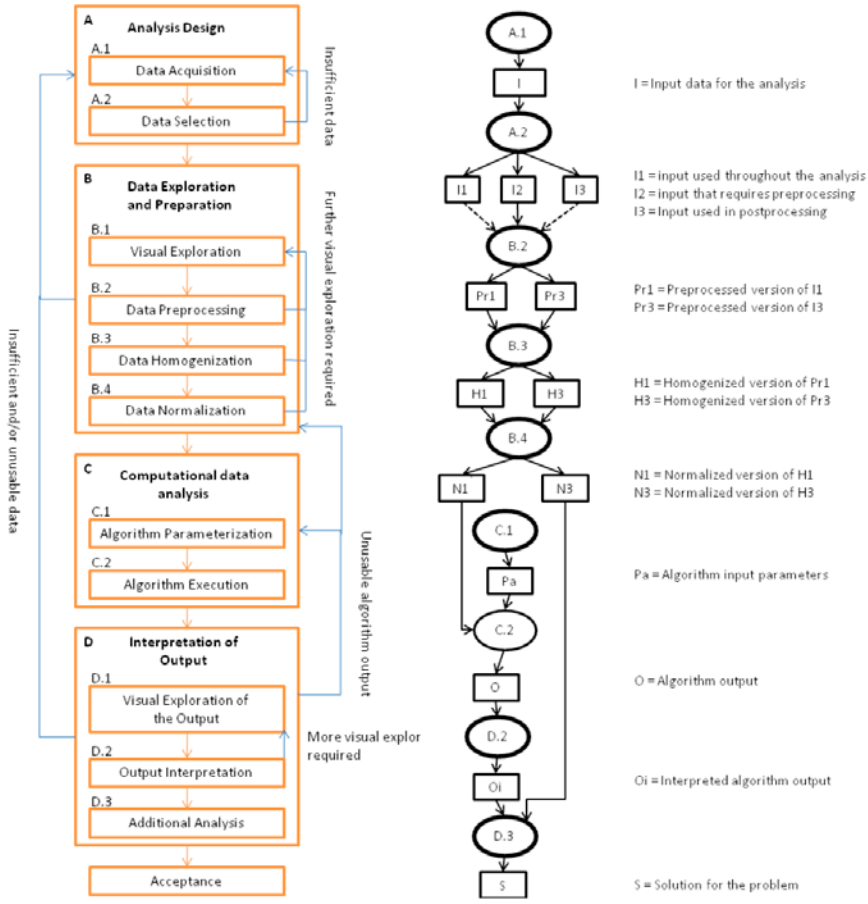

Figure 1. *Overview of the analysis process and the data flow. The process, on the left-hand side of the figure, consists of four phases divided into subphases. The data flow, on the right-hand side, shows how the input data is modified in various parts of the process.*

phase, since in this step the user runs one of the algorithmic methods. The output of the algorithm needs to be visualized and interpreted before the user can decide whether he or she needs to move back to a previous phase. This is done in phase D of the analysis process.

On the right-hand side of Figure 1 is the data flow of the process. The ovals stand for subphases and rectangles for input/output sets. The bolded ovals stand for subphases where the user's expert knowledge is used to guide the analysis process. A set is used as input if an arrow goes from the rectangle depicting the sect to a phase of the process. A set is the output of a given phase, if there is an arrow from the phase to the set. There are dashed arrows from sets I1 and I3 to the phase B.2 of the process. This is used to depict the fact that while I1 and I3 are not actually used as input for the phase, the phase still affects the contents of these sets, and the output sets Pr1 and Pr3 are distinct from I1 and I3. Far left in the Figure are explanations for the various data sets and what they are used for.

We assume that the user who employs the analysis process has the required expertise to solve the problem at hand. Since the process itself does not contain any knowledge or values, the user needs to have sufficient expertise to add these into the process.

There are two kinds of knowledge that is required in order to solve the problem. First is domain knowledge: knowledge about the problem at hand, the factors that affect it, and how these factors affect one other. The second is GIS knowledge: knowledge about how to use and analyze spatial data and how to use spatial data in problem solving. A user who has sufficient domain knowledge is called a domain expert, and a user who has sufficient GIS knowledge is called a GIS expert. It is possible that a single user is both domain and GIS expert, but in this work we will discuss them as separate people. This makes it easier to distinguish between cases where domain expertise is required from cases where GIS expertise is required. (Krisp 2006)

3.1 Phase A: Analysis design

In phase A, *analysis design*, the user acquires the geographic data sets that are used as input for the rest of the process (phase A.1) and categorizes the acquired data sets according to how they are used in the analysis (phase A.2). As shown in the left-hand side of Figure 1, the output of phase A.1 is set of data layers *I*, which contains all input data layers used in the analysis process. In phase A.2 this set is divided into three subsets I1, I2, and I3. Set I1 contains all data layers that are used through the data analysis process, and I3 contains data layers that are not suitable for computational analysis in phase C. Set I2 contains all data layers that require preprocessing.

Here, the user's expert knowledge comes into play for the first time. In phase A.1 domain knowledge is required for analyzing the problem at hand, and deciding what data and information is required for solving the problem. GIS knowledge is required in analyzing the possible input data, and finding which of these are available as spatial datasets, and where such spatial data can be found.

In phase A.2 the user must be able to separate datasets that can be used in computational analysis from data sets that are best included in the additional analysis phase. They must also be able to see which of the data sets require preprocessing before being used in the analysis. This requires both domain knowledge to know how a given input data layer affects the problem at hand, and GIS knowledge in order to know how the input data can be used in the analysis.

3.2 Phase B: Data Exploration and Preparation

In this phase of the process, the user familiarizes themselves with the details of the input data, and can modify it in various ways.

In phase B.1, *Visual Exploration*, the user can use various visualizations to explore the input data layers and familiarize themselves with the details of the input. In phase B.2, *Data Preprocessing*, the user transforms the input data layers contained in set *I2* so that they can be used in further phases of the process. In phase B.3, *Data Homogenization,* the user transforms the input layers so that they

all use the same coordinate projection and resolution, and thus can be overlaid further in the analysis process. Finally, in phase B.4, *Data Normalization,* the user transforms the data layers into normalized format, where the data values of various layers can be compared, and the data can be used as input for phase C of the process.

In phase B.1 GIS knowledge is used to understand the contents of the various visualizations used for exploring the data, and drawing inferences from it. This process also requires domain knowledge in order to understand how the different input layers affect the problem. In phase B.2 both GIS knowledge and domain knowledge are required in order to select appropriate preprocessing for the input layers. Domain knowledge is required in order to know what the data layers need to represent after the preprocessing, and GIS knowledge is required in order to know how to preprocess the layers. In phase B.3 GIS knowledge is required in order to select appropriate coordinate projection and resolution. In phase B.4 domain knowledge is required in order to know how each data layer independently affects the problem, and thus how that particular layer should be normalized.

3.3 Phase C: Computational Data Analysis

In this phase of the process, the user selects an analysis method and feeds the input data to the method. The output of the method is then interpreted in phase D. This phase consists of two subphases. In phase C.1, *Algorithm Parameterization,* the user selects an appropriate algorithm and parameters for it. In phase C.2, *Algorithm Execution,* the user runs the selected algorithm using the parameters selected in phase C.1 and the normalized input data created in phase B.4.

This phase of the process requires mainly knowledge about the algorithms incorporated into the process and thus cannot easily be categorized either as domain or GIS knowledge. In this context it is perhaps closer to GIS knowledge, since it concerns the tools used for solving the problem, and not knowledge about the various aspects of the problem itself.

3.4 Phase D: Interpretation of the Output

The usefulness of the output is reviewed in phase D of the analysis process, the *interpretation of the* results. In this phase, the user uses different visualizations to explore the output of the algorithmic data analysis method, interprets and postprocesses the results, and decides whether the output is acceptable. This phase consists of three subphases.

In phase D.1, *Visual Exploration of the output,* the user applies various visualizations to familiarize themselves with the algorithm output O. In phase D.2, *Output Interpretation,* the user interprets the how the various parts of output O fit the problem at hand and categorizes each part into one of the desired output classes. Finally, in phase D.3, *Additional Analysis,* the user adds the input data layers that were not included in the computational analysis to the solution. This is followed by the final phase of the process, *Acceptance*, where the user accepts the produced solution as a solution for the problem, or rejects it and continues the process from an earlier phase.

This phase of the process requires both GIS and domain knowledge. Phase D.1 requires both types of knowledge as the user explores the algorithm output. In phase D.2 domain knowledge is required for interpreting how well each part of the output suits the activity that is being analyzed and what sort of suitability value it should be given. Phase D.3 requires both GIS and domain knowledge in order to include the input layers of *N3* to the overall solution. Similarly the final acceptance or rejection of the solution requires both GIS and domain knowledge.

4 The Case: Off-Road Mobility

We tested out the analysis process model described in Section 3 in a real analysis situation. The example problem was off-road mobility – or cross-country mobility – for a vehicle, which is the ability of a specific vehicle to travel outside the established road network. It is typically depicted using *a mobility map*. A mobility map is a type of *cost surface*, where the value of each pixel represents the amount of resources required for specific activity (movement) at that location. In this experiment, mobility was divided into three categories: GO, GO SLOW, and NO GO. The names for the categories come from military use, and they represent areas that are well-suited (have high mobility), suitable (allow for movement), and unsuitable (cannot be traversed).

A three-value mobility map is typically used for finding out where it is possible to move, and has limited value in further analysis due to the small number of attribute values. A map with a wider range of attribute values, on the other hand, can allow for further analysis. For example, if the map is such that the value of each pixel represents the amount of time it takes to travel across the area covered by the pixel, it can be used for calculating the fastest path. Such analysis cannot be done using only three mobility categories, since each category contains a broad range of travel times.

The example described here follows the process model described in Section 3, and therefore the names introduced in Section 3 are used for the phases of the process.

4.1 Analysis Design

The first step in a spatial analysis process is to gather appropriate input data, and categorize it. The factors that affect mobility are the soil type, the amount and type of vegetation, the degree of slopes, roads, and buildings. Roads typically offer better mobility than any off-road situation, and buildings prevent mobility. There are often also factors that are part of a specific environment. For example, in northern latitudes snow and frost are important factors during the winter.

The data layers may require preprocessing, and are used in different ways during the analysis process. The effect that roads and buildings have on mobility is not influenced by other input data layers. Therefore these two data sets should be included at the end of the analysis process. The other data layers, on the other hand, need to be combined to know the overall effect they have on mobility, and thus are best used throughout the analysis.

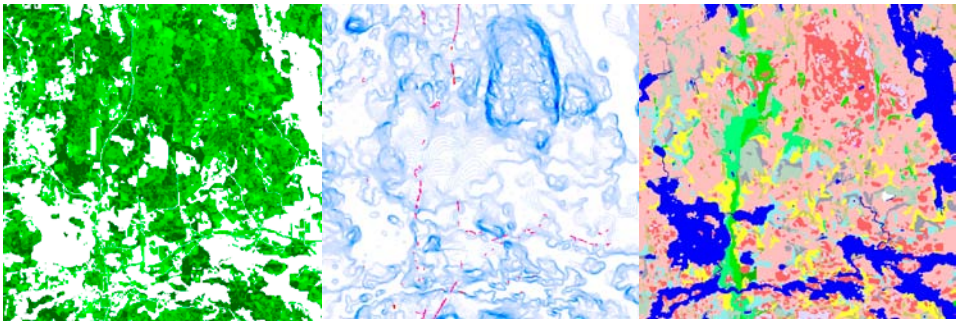

Figure 2. *Vegetation, slope, and soil raster layers. In the vegetation layer, the deeper the green, the more vegetation there is. In the slope layer, the deeper the blue, the steeper the slope. Extremely steep slopes are represented by using red. Different colors in the soil layer stand for different soil types, and blue stands for water.*

In this case several data layers require preprocessing before they can be used in further analysis. For example, elevation data needs to be transformed into raster data, and vector data needs to be rasterized. Thus, using the data flow categorization shown in Figure 1, the output for phase A is I1 = {vegetation, soil, elevation}, I2 = {elevation, buildings, road}, I3={buildings, road}.

4.2 Data Exploration and Preparation

After the data has been gathered and classified, the next step in the process is to explore the data sets and then modify them to be comparable, and usable as input for a computational method. The first modification done is preprocessing, where the data is transformed into a form usable in the rest of the process. For example, elevation data must be transformed into slope data. Figure 2 contains example snapshots of how the vegetation, slope and soil data types look after preprocessing. The actual examples used here are from data preprocessed by the Finnish Defense Forces.

After preprocessing, the next step is to homogenize all the data sets. In this case, the datasets are already available at least in 1:20,000 resolution, and most use the proper coordinate system. After homogenization, the datasets need to be normalized. Here, the two data sets need to be normalized differently. The data, which will be used in phase C, need to be normalized to some closed interval in order to make the data sets comparable. The data, which is added at the end of the process, need to be normalized using the output mobility categories.

The first type of normalization can be the interval from 0 to 1. Here 0 stands for no mobility and 1 for perfect mobility. For example, in the slope layer flat ground would give best mobility, and mobility decreases as the slope increases. How steep slopes can be traversed depends on the vehicle: an off-road vehicle can typically clear steeper slopes than a normal ground car.

The second type of normalization would give the data values corresponding to GO, GO SLOW, or NO GO. In this case the categorization is easy. Roads

give good mobility, and would thus correspond to GO, whereas buildings prevent mobility and would therefore be NO GO. The output for phase B is datasets N1 and N3, which are normalized versions of H1 and H3. Thus, N1={vegetation, slope, soil}, and N3={roads, buildings}.

4.3 Computational Data Analysis

When the data has been properly normalized, the user apply a computational analysis method in order to gain more information from the data. The method used in this case is k-means clustering (MacQueen 1967), with k set 10, giving an output with ten clusters. Typically, the best number of clusters is something that need to determined experimentally.

As output k-means gives 10 clusters, the cluster centers and members, as well as a map showing the geographic distribution of the clusters. Or, more formally, output set O = {k0,…, k9, map}.

4.4 Interpretation of the Output

The output of the computational method needs to be interpreted by the user. This requires visualizations for showing clustering results. One good multivariate data visualization is the parallel coordinates plot (PCP), which is used to show n-dimensional data points with polylines that have vertices on the parallel axes. In this case, the data set is 4-dimensional points, where each point contains a value from each of the three input layers and a cluster number. A map view can be used to show the geographic distribution of each cluster.

In order for the user to explore the data, the two data views must be linked; when the user highlights an area from the PCP, corresponding points from the map are also highlighted, and vice versa. Figure 3 contains an example of linked data views containing a map and a PCP visualization. One of the clusters has been highlighted on the map, and the PCP shows the data values in the cluster.

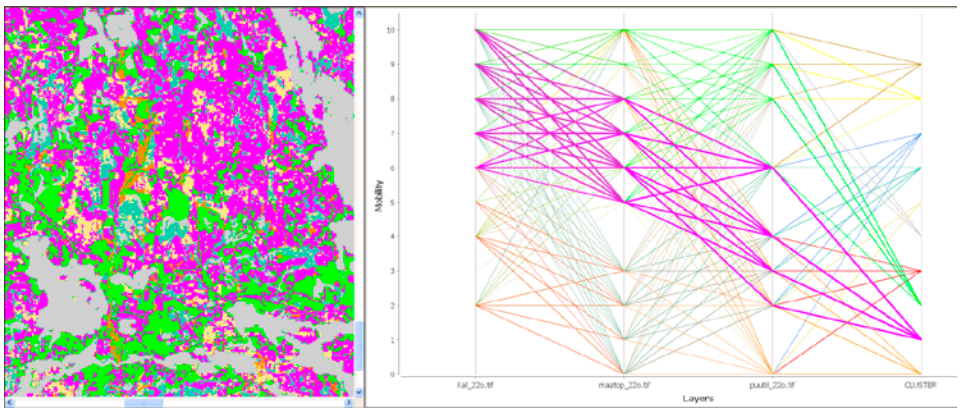

Figure 3. *Linked map and parallel coordinates plot view of k-means results. One of the clusters has been highlighted in both views.*

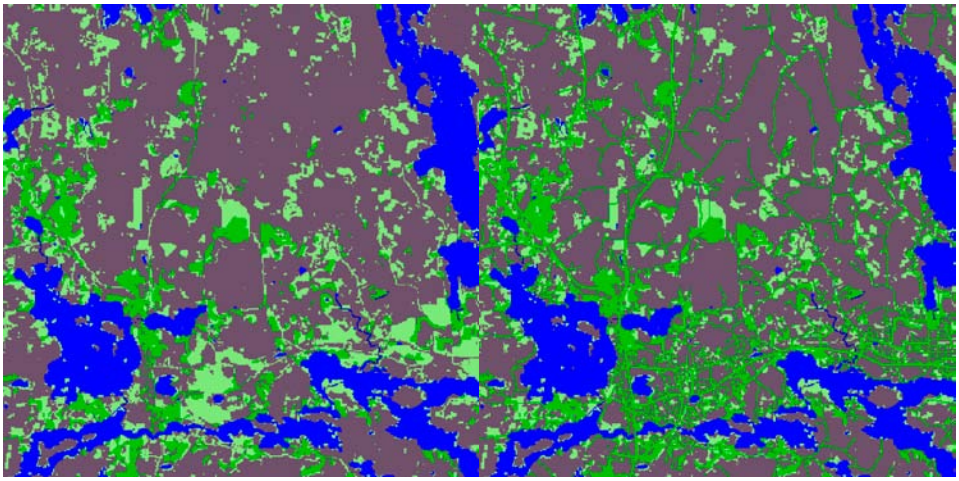

Figure 4. *Mobility map before (on the left) and after (on the right) the inclusion of road and urban data layers.*

After familiarizing themselves with the output, the user can interpret it by giving each cluster a mobility value. For this, user needs to view both the attribute values contained in a cluster and the geographic distribution of the cluster. After each cluster has been given a mobility value, the cluster map can be turned into a mobility map. Then, the user can add the road and building data to the mobility map. Since this data has already been normalized to using the three mobility categories, this phase consists of overlaying the two layers over the map. Figure 4 shows a mobility map before and after the inclusion of road and building information. Some road-like features can be seen on the map before the inclusion of roads due to the good quality of the input data. The input data shows, for example, areas where vegetation has been cleared to make ways for roads, and this is reflected in the clustering results.

5 Prototype Application and Results

In order to validate the usefulness of our approach, we created a prototype application for analyzing off-road mobility in Finland. The prototype was implemented using the ESRI ArcObjects GIS framework, as well as freely available jCharts and JFreeChart graphics libraries. The prototype covers phases B.4 through D.2 of the analysis process, since these are the parts of the analysis process that are not included in a typical desktop GIS environment. For the rest of the process we used the ESRI AcrGIS environment. Included in the prototype there are two computational methods: k-means and DBSCAN clustering. These two were selected for the initial prototype since they are simple to understand and teach to users, simple to implement, and have been used for solving many GIS problems.

We know that with more complex clustering methods, which take into account the spatial distribution of the data, it is possible to get better results. However, the focus of this research is not in algorithms, and therefore we felt that

using simple, well-known methods is sufficient for our purposes. Furthermore, it makes it easier to maintain the transparency of the data analysis process, since the algorithms are well-known and sufficiently simple to easily examine and verify.

Our results were then validated by comparing them to mobility maps created by the Finnish Defense Forces. The Defense Forces have a detailed mobility model for Finnish terrain which has been tested using both expert evaluation and in actual field experiments. In the following, we will refer to this model as the FDF model. The original FDF model has seven mobility classes. For comparison purposes, we generalized them to three classes standing for GO, GO SLOW, and NO GO areas.

We performed an independent off-road mobility analysis using the methods presented here with the input data used for creating the FDF model. The normalizations we used were based on the work of the Defense Forces, with some modifications. All data were normalized between 0 and 1 with one-digit precision. Zero stood for no mobility and one for maximum mobility. The exact mobility map we were trying to create was mobility for Patria Pasi armored personnel carrier in the summer. The area was part of central Finland which had both wilderness and urban areas. The area covered was 80 km × 80 km in size, and a pixel was 25 m × 25 m, making the picture 3,200 × 3,200 pixels in size. Here we report the k-means and DBSCAN interpretations that best correspond to the FDF model. Both interpretations include roads and urban areas. In order to simulate ad hoc analysis, all roads were assigned good mobility (GO) and urban areas were assumed to prevent mobility (NO GO). Thus, we assumed that no data about road or urban area types were available. In reality, the urban area dataset included, for example, parking areas and airstrips, which offer good mobility. In places where the two datasets overlapped, roads were given precedence over urban areas.

5.1 Visual Comparison

We did an initial investigation of the datasets by simply comparing the mobility maps visually. We noticed that areas of good mobility (GO) seemed to be quite similar in the k-means and the model data sets. The data set created using DBSCAN clustering apparently had more areas of good mobility than either k-means or model sets. A lot of area that contained good mobility (GO) in the DBSCAN interpretation was either fair (GO SLOW) or even bad mobility (NO GO) in the k-means and model maps.

The difference in areas of good mobility between the k-means interpretation and the DBSCAN interpretation was due to differences in the algorithm output. Both k-means and DBSCAN produced one cluster that can be interpreted as good mobility. The clusters were, however, quite different and thus covered different parts of the map. Figure 5 contains PCP visualizations of the clusters. As can be seen from the figure, the k-means clustering result is clearly an area of good mobility, since all values are at least 0.6. The DBSCAN cluster, on the other hand, contains a number of data elements where values are 0.4 in the slope layer and

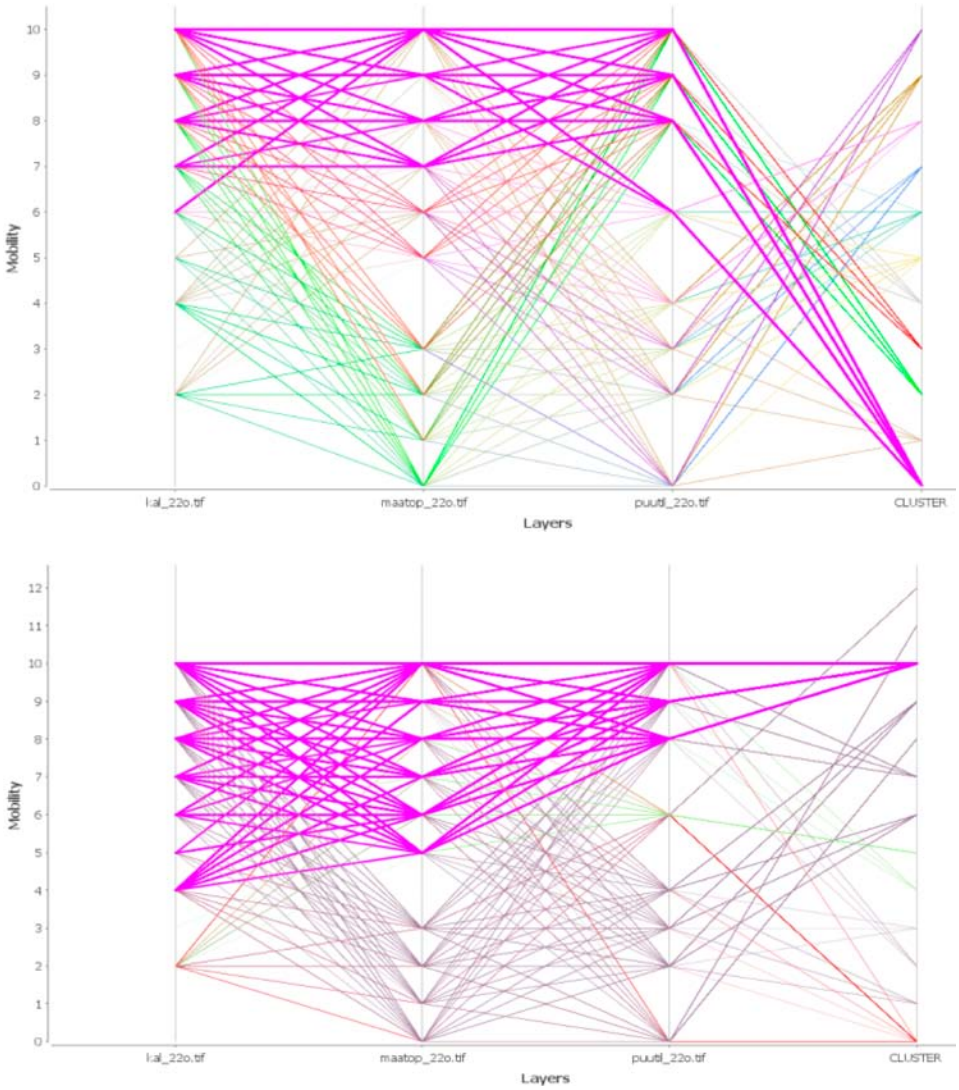

Figure 5. *Parallel coordinates plot representations of clusters of good mobility in k-means and DBSCAN results. Note the difference in the vertical scales, since the DBSCAN produced a total of 13 clusters. The data axes are, from left to right: slope degree, soil type, vegetation, and cluster number.*

0.5 in the soil layer and thus might not undoubtedly be good mobility. Similarly, there were differences between the DBSCAN and k-means clusters that could be interpreted as fair mobility (GO SLOW).

5.2 Misclassification Analysis

The dataset used in this study contained over 10 million data elements distributed between four different values: unclassified, go, go slow, and no go. Unclassified

elements are those for which a mobility value cannot be assigned by a given method. Table 1 summarizes the data distribution in all three clustering results and the model data set. As can be seen from the table, the k-means interpretation has no unclassified data items, since k-means assigns each data item into a cluster. DBSCAN is capable of detecting noise and outliers and includes unclassified pixels.

Table 1. *Summary of experimental data. For each dataset, both the absolute and relative numbers of elements for each of four data values are given.*

	Model data	%	DBSCAN	%	k-means	%
unclassified	26,008	0.002	5,726	0.0006	0	0
Go	871,825	0.085	1,962,356	0.191	1,013,211	0.098
go slow	1,052,816	0.102	744,222	0.072	1,273,061	0.124
no go	8,289,109	0.809	7,527,454	0.735	7,953,486	0.776
Total	10,239,758	1	10,239,758	1	10,239,758	1

To further investigate how well our interpretations fit the model data, we created misclassification matrices for our interpretations. The misclassification matrices were created comparing each map against the Defense Forces' mobility map. We wanted to investigate how big the differences actually were between our interpretations and the existing mobility maps. Misclassification matrices for the two datasets can be found in Tables 2.1 and 2.2. The tables show us the misclassification between each class in the model data set and the interpretation. At the bottom of each table we have also included the kappa index for the misclassification. The kappa index has values between zero and one, and it shows how much better the classification is compared to a totally random distribution of data values. Zero corresponds to totally random distribution and one to a perfect match between classifications.

Each row in the tables shows how one data value in the clustering result is divided between data values in the model data set. Both clustering results and the model data set have four possible data values. In addition to the different mobility values, there are data elements for which a mobility value could not be given. Such items are marked as unclassified in the tables. Since k-means clustering places all data elements into clusters, there are no unclassified elements in the k-means clustering results. DBSCAN is capable of spotting noise and outliers, and thus can create unclassified elements.

Table 2.1. *Misclassification matrix between DBSCAN clustering and model data set.*

Model DBSCAN	Unclassified	Go	Go slow	No go	Total
Unclassified	0	0	71	5,655	5,726
Go	13,585	849,835	891,514	207,422	1,962,356
Go slow	0	0	69,120	675,102	744,222
No go	12,423	21,990	92,111	7,400,930	7,527,454
Total	26,008	871,825	1,052,816	8,289,109	10,239,758
Kappa					0,51

Table 2.2. *Misclassification matrix between k-means clustering results and model data. set*

Model K-means	Unclassified	Go	Go slow	No go	Total
Unclassified	0	0	0	0	0
Go	13,585	808,775	190,816	35	1,013,211
Go slow	0	59,148	740,447	473,466	1,273,061
No go	12,423	3,902	121,553	7,815,608	7,953,486
Total	26,008	871,825	1 052,816	8,289,109	10,239,758
Kappa					0,76

As can be seen from the two tables, there are significant differences between the cluster interpretations. The first big difference is in the distribution of data elements assigned good mobility in the DBSCAN and k-means interpretations. In DBSCAN, the total number of data elements assigned good mobility was much higher than in the k-means interpretation. In DBSCAN approximately 19% of the total area demonstrates good mobility, whereas in k-means approximately 10% of the area contains good mobility. For reference, in the model data set, approximately 8.5% of the area contains good mobility. A more significant difference than the sizes of GO areas is their spatial distribution. Both DBSCAN and k-means interpretations cover most of the model's GO area, as well as some of the model's unclassified pixels. The DBSCAN interpretation, however, also classifies a large amount of area as GO that, in the model, is either GO SLOW or even NO GO. The k-means interpretation, on the other hand, contains significantly less misclassified GO area, and practically all of it is GO SLOW in the model. In the DBSCAN interpretation, the significant amount of GO area that is NO GO in the model is problematic since it represents a major difference between the two solutions. Furthermore, the GO SLOW area in the DBSCAN interpretation is mostly NO GO in the model interpretation. Thus, the k-means interpretation clearly corresponds better to the model than the DBSCAN interpretation.

The NO GO areas in the DBSCAN and k-means interpretations correspond rather well to the NO GO areas in the model. In the two interpretations the amount of misclassified area is less than 2% of the total data set. For example, there are less than 2% of the total dataset which has been marked as NO GO in either interpretation and is marked GO SLOW or GO in the model. There are also some data elements that are GO in the model and NO GO in the interpretations, but the amount of such data is rather small in both interpretations.

The kappa indices show us that the k-means output fits the model rather well. The Kappa index of 0.76 indicates that the classification fits the model very well. The kappa of the DBSCAN is smaller, at 0.51. However, the k-means interpretation clearly shows that by using clustering it is possible to gain results similar to the model.

5.3 Expert Opinion

In order to further validate the usefulness and validity of our approach, we showed both the DBSCAN and k-means clustering results to a Defense Forces expert. The

expert classified both clustering results and gave further comments on the results. We had given the same interpretation to most of the clusters as the expert did, but there were also some differences. The biggest find in the expert evaluation was, however, that according to the expert's opinion several clusters should have been split according to one of the data axes. For example, according to the expert, the DBSCAN GO cluster should be split according to the slope axis, and areas where the slope mobility is at least 0.7 should be GO and the rest GO SLOW.

There were also some other clusters which should be split, according to the expert's opinion, and a few clusters which would have required more detailed analysis. The expert would have wanted, for example, to be able to see the original soil types included in some clusters, or be able to look at the single data vectors in the PCP in more detail. In our rather simple prototype, these user interface options had not been implemented. Thus, the expert was unable to give a few clusters an exact classification.

However, for the most part the expert's opinion corresponded to our interpretation. Furthermore, the expert suggested several improvements to the prototype, the most important of which is the ability to split clusters according to a given data axis.

6 Discussion and Conclusions

The results of this work clearly indicate that it is possible to use an exploratory, user-controlled, and interactive approach together with clustering to gain good analysis results for the cross-country mobility problem. By using k-means clustering we were able to create a mobility map that had corresponded well to the result created using traditional model-based approaches. For this particular analysis, the results of the DBSCAN clustering did conform to the model as well as k-means. Both algorithms were used in an interactive and iterative process, where the control of the process was in the hands of the user, and in both cases the algorithms required several runs before the results described here were achieved. On each run, the user parameterized the algorithm and interpreted the results.

Since this work discusses the results from the cross-country mobility analysis, we have used the mobility terms NO GO, GO SLOW, and GO in the text. These are problem-specific terms, but can easily be generalized. NO GO describes unsuitable or non-fit areas, GO SLOW corresponds to areas that are fairly suitable, and GO to areas that are well suited or best-fit to the problem at hand. As many problems can be abstracted into dividing the possible sites into these three categories, the analysis process can be considered generic in this sense.

The methods created here have been be utilized without using any values or knowledge built into the process itself. Instead, the process is user-controlled, and the expert user inserts their knowledge into the process. The knowledge insertion starts in the first phase of the process, with the selection of inputs. Perhaps more important from our point of view is, however, the knowledge inserted in the data preparation and output interpretation phases of the process. While preparing the data for the computational analysis methods, the user needs to insert their

knowledge in the form of supplying the normalization parameters for the data. The normalization then has a great influence on how the data are clustered by the analysis methods. Also of great importance is the user's knowledge in interpreting the output and deciding whether the output gained is usable for the situation at hand.

The two clustering methods used in this work seem to be sufficient for the analysis process. They are, for example, very time-efficient. Most of the time involved in running a clustering algorithm is taken by preprocessing and postprocessing the data. The execution of the actual clustering typically takes only a fraction of the whole running time. This is mostly because clustering methods are run using distinct data vectors. Thus, before running a clustering algorithm, the system filters the distinct data vectors from the input, and after the clustering has been created a new cluster map must be expanded from the result. The number of distinct data vectors is typically much smaller than the total number of pixels in the input, and thus the filtering and expansion steps are much more time-intensive tasks than analyzing the data set. The output of the simple clustering algorithms does not always seem to be sufficient for gaining good results. This can be seen, for example, in the DBSCAN clustering result. The cluster, which contains good mobility, contains also pixels that cannot be considered good mobility by any measure. Similarly, the fair mobility cluster in DBSCAN contains mostly area that is NO GO in the model.

In solving the cross-country mobility problem the use of clustering, linked views, and interactivity gives the user a much more detailed view of the problem at hand than the traditional methods used by the Defense Forces. By dividing the data into similar subsections, clustering can reveal distinct clusters in the data. The user can then explore these by using several visualizations at the same time, and thus can gain a detailed picture of how each cluster affects mobility. Thus, the user is not limited to some preset rules about how to divide the area into different mobility classes, but can take into account the details of the current situation. Moreover, using clustering the user can see how certain areas hinder mobility. For example, if a cluster contains areas that would offer good mobility if not for the extensive amount of vegetation there, the user can see this from the visualization. Thus, if no other routes are possible, the vegetation can be cleared, improving the mobility of the area, or a path through it can be created. Finding such areas or making such decisions would be impossible using the traditional methods, since they do not preserve information about what hinders mobility in a given area.

In this work, we have described a flexible, user-controlled, and values-free analysis process, primarily aimed at international civilian crisis management, and have used the process to solve the cross-country mobility problem. The process combines visual analytical approaches with interactivity and computational methods. The visualizations used in the system include parallel coordinate plots and map views, which are used to interpret the results of the clustering. Two clustering algorithms are currently implemented: k-means and DBSCAN. A prototype system has been constructed, and tested by analyzing the cross-country mobility problem for vehicles. Compared to previously used methods, the use of

linked visualizations and clustering reveals previously unseen information in the data, and enables more flexible and involved decision making to take place. The results indicate that our methodology is sound and that the process can be used to solve these problems.

6.1 Future Work

We have successfully used the method to analyze one problem, and have started work on using it in solving other problems. The current prototype we have constructed is of rather limited functionality, but it is still under active development. In the future, we are going to include new interactive functionalities, including the ability to divide clusters into smaller ones, and expand the analysis by including methods that explicitly take into account the neighborhood of each element. The current version handles each raster pixel separately, without taking the neighborhood into account. This prevents us from taking into account, for example, the proximity of roads or other interesting objects, viewsheds, or cover created by terrain features. Furthermore, we are going to apply the methodology to solving new problems and thus demonstrate the generality of our approach. Additionally, we are going to perform large-scale user testing of the system in the future.

In our opinion our interactive approach to data analysis has great potential. We have already used it successfully in solving cross-country mobility, which is a common problem faced in international crisis management. We have made preliminary studies on applying the same methodology to other problems, such as finding suitable locations for different types of facilities (e.g., supply depots, field hospitals, artillery battalions, etc.) and analyzing locations for communications link masts. So far, we do not have any definitive results to show for these applications, but initial results look promising.

Bibliography

Andrews, D. F. (1972), 'Plots of high-dimensional data', *Biometrics* 29, 125–136.

Andrienko, G. & Andrienko, N. (2001), Constructing parallel coordinates plot for problem solving, *in* '1st International Symposium on Smart Graphics', New York, USA, pp. 9–14.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., Jern, M., Kraak, M.-J., Schumann, H. & Tominski, C. (2010), 'Space, time and visual analytics', *International Journal of Geographical Information Science* 24(10), 1577–1600. http://www.tandfonline.com/doi/abs/10.1080/13658816.2010.508043

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M.-J., MacEachren, A. & Wrobel, S. (2007), 'Geovisual analytics for spatial decision support: Setting the research agenda', *International Journal of Geographical Information Science* 21(8), 839–857. http://www.tandfonline.com/doi/abs/10.1080/13658810701349011

Berkhin, P. (2002), Survey of clustering data mining techniques, Technical report, Accrue Software, San Jose, CA. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.3739

Bertin, J. (1981), *Graphics and Graphic Information Processing*, de Gruyter.

Chambers, J., Cleveland, W., Kleiner, B. & Tukey, P. (1983), *Graphical Methods for Data Analysis*, Wadsworth.

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* 35(1), 33–50.

Chernoff, H. (1973), 'The use of faces to represent points in k-dimensional space graphically', *Journal of the American Statistical Association* 68(342), 361–368.

Demšar, U. (2006), Data mining of geospatial data: combining visual and automatic methods, PhD thesis, Royal Institute of Technology.

Demšar, U., Špatenková, O. & Virrantaus, K. (2008), 'Identifying critical locations in a spatial network with graph theory', *Transactions in GIS* 12(1), 61–82.

Edsall, R. (2003), 'Design and usability of an enhanced geographic information system for exploration of multivariate health statistics', *The Professional Geographer* 55(2), 146–160.

Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise, *in* 'Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)', AAAI Press, pp. 226–231.

Gorsevski, P., Jankowski, P. & Gessler, P. E. (2005), 'Spatial prediction of landslide hazard using fuzzy k-means and Dempster-Shafer theory', *Transactions in GIS* 9(4), 455–474.

Guo, D. (2008), 'Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap)', *International Journal of Geographical Information Science* 22(7), 801–823.

Guo, D., Gahegan, M., MacEachren, A. M. & Zhuo, B. (2005), 'Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach', *Cartography and Geographic Information Science* 32(2), 113–132.

Horttanainen, P. & Virrantaus, K. (2004), Uncertainty evaluation by simulation and visualization, *in* 'Geoinformatics 2004'.

Inselberg, A. (1985), 'The plane with parallel coordinates', *The Visual Computer* 1, 69–97.

Jiang, B. & Harrie, L. (2004), 'Selection of streets from a network using self-organizing maps', *Transactions in GIS* 8(3), 335–350.

Keim, D. A. & Kriegel, H.-P. (1996), 'Visualization techniques for mining large databases: A comparison', *IEEE Transactions on Knowledge and Data Engineering* 8(6), 923–938.

Keim, D., Kohlhammer, J., Ellis, G. & Mansmann, F., eds (2010), *Mastering the Information Age – Solving Problems with Visual Analytics*, Eurographics Association.

Krisp, J. (2006), Geovisualization and Knowledge Discovery for Decision-making in Ecological Network Planning, PhD thesis, Helsinki University of Technology.

Liao, K. & Guo, D. (2008), 'A clustering-based approach to the capacitated facility location problem', *Transactions in GIS* 12(3), 323–339.

MacEachren, A. M., Wachowicz, M., Edsall, R., Haugh, D. & Masters, R. (1999), 'Constructing knowledge from multivariate spatiotemporal data: integrating geographical

visualization with knowledge discovery in database methods', *International Journal of Geographical Information Science* 13(4), 311–334.

MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, *in* 'Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability', University of California Press, pp. 281–297.

Mahalanobis, P. C. (1936), 'On generalized distance in statistics', *Proceedings of the National Institute of Sciences in India* 2, 49–55.

Orava, E. (1997), Terrain analysis for military use, Master's thesis, Helsinki University of Technology.

O'Sullivan, D. & Unwin, D. J. (2003), *Geographic Information Analysis*, Wiley, chapter Putting Maps Together: Map Overlay, pp. 284–314.

Pei, T., Zhou, C., ZHu, A.-X., Li, B. & Qin, C. (2010), 'Windowed nearest neighbour method for mining spatio-temporal clusters in the presence of noise', *International Journal of Geographical Information Science* 24(6), 925–948.

Sander, J., Ester, M., Kriegel, H. & Xu, X. (1998), 'Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications', *Data Mining and Knowledge Discovery* 2(2), 169–194.

Seppänen, H. & Virrantaus, K. (n.d.), 'The role of GIS methods in crisis and disaster management', *International Journal of Digital Earth*. Accepted for publication.

Shaw, S.-L., Yu, H. & Bombom, L. S. (2008), 'A space-time GIS approach to exploring large individual-based spatiotemporal datasets', *Transactions in GIS* 12, 425–441.

Shoshany, M., Even-Paz, A. & Bekhor, S. (2007), 'Evolution of clusters in dynamic point patterns: with a case study of ants' simulation', *International Journal of Geographical Information Science* 21(7), 777–797.

Sips, M., Schneidewind, J. & Keim, D. A. (2007), 'Highlighting space-time patterns: Effective visual encodings for interactive decision-making', *International Journal of Geographical Information Science* 21(8), 879–893.

Stefanakis, E. (2007), 'NET-DBSCAN: clustering the nodes of a dynamic linear network', *International Journal of Geographical Information Science* 21(4), 427–442.

Thomas, J. J. & Cook, C. A., eds (2005), *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, National Visualization and Analytics Center.

Vesterinen, S. (2008), Shift shared information frameworks and technology concept draft 0.5, Technical report, Edita Prima OY.

Virrantaus, K. & Horttanainen, P. (2004), Developing a knowledge-based spatial uncertainty model, *in* 'Proceedings of the 3rd International Symposium on Spatial Data Quality'.

Zhang, J. & Goodchild, M. F. (2002), *Uncertainty in Geographic Information*, Taylor and Francis.

Zhang, Z. & Virrantaus, K. (2010), Analysis of vulnerability of road networks on the basis of graph topology and related attribute information, *in* 'In Proceedings of 2nd International Conference on Intelligent Decision Technologies'.