❄

Evaluating the appropriateness of visually combining quantitative data representations with 3D desktop virtual environments using mixed methods

Susanne Barbara Bleisch

Thesis Submission for Admission to degree of
Doctor of Philosophy
in Geographic Information Science

City University London
Information Science

Mai 2011

Contents

Each of the chapters (1-5) starts with a box like this one where the whole chapter is summarised in an informative abstract. Reading through all these boxes should give you a quick overview of the whole thesis. Additionally, these summaries serve as overview of each chapter by providing links (by section numbers) to the respective sections containing detailed information.

List of Tables

List of Figures

Acknowledgements

I am most grateful to:

Jason Dykes for accepting to supervise me as a distance part-time PhD student, for the inspiring discussions face-to-face and via Skype, for his invaluable feedback on my thinking and writing.

Stephan Nebiker for supporting the idea of writing a dissertation from the very beginning. He provided encouragement and stimulating discussions and always seems to be one step ahead with thinking and creating new ideas. Thank you for offering a work environment that allowed writing a dissertation at FHNW.

Reinhard Gottwald and the Institute of Geomatics Engineering at FHNW for my employment and supporting my endeavour financially. Participation at national and international workshops and conferences and thus many interesting discussions with other researchers would not have been possible without this support.

Menno-Jan Kraak and Aidan Slingsby for acting as independent reviewers of this dissertation.

All the participants of the experiments conducted in the course of this research. Without you taking the time, working through the experiments and providing feedback, this research would not have been possible. Special thanks go to the case study data experts Peter Mahler, Barbara Piatti and Seraina Campell for providing their data, taking the time to participate and provide feedback.

All my colleagues both at FHNW and City University for support, interesting discussions on various topics and answering questions. Special thanks go to Hannes Eugster, Joel Burkhard, Kevin Flückiger, Adrian Annen, Stephan Schütz, Natalie Lack, Martin Christen and Andreas Barmettler at FHNW and Jonathan Raper, Jo Wood, David Lloyd, Iain Dillingham, Naz Khalili-Shavarini and Lian-Chee Koh at City University.

Ruedi Haller and the Swiss National park for providing the deer tracking and background data sets (swisstopo 2010).

Sarah Castle for carefully proofreading the final draft of this dissertation.

The many researchers, known and unknown, who took the time to read drafts and provide feedback on articles and conference papers and who published their research so that I could learn from them.

The developers of LaTeX. It is the most stubborn but, more importantly, also most reliable piece of software I have ever worked with.

Last but not least, my family for their support and encouragement during all my life.

. . . and especially Urs for sustained encouragement and distraction likewise.

Olten, Switzerland, March 2011
Susanne Bleisch

Spelling convention

There is an American spelling adopted throughout the thesis to align with the commonly used name for a research area. This is 'geovisualization' for 'geovisualisation'.

Declaration

"I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement."

Abstract

The integration of various data sets into desktop based 3D virtual environments, such as the virtual globe Google Earth, is quickly achieved with today's technological options. Nevertheless, we know little about the appropriateness of such representations. A number of research studies have looked at different aspects of 3D virtual environments, in particular interaction and navigation, but rarely at the use of virtual environments for data analysis. The visual combination of quantitative data with the three-dimensional virtual equivalent of the natural environment where a data set was collected may help the analysis of such data sets in regard to altitude and landform. Data sets demonstrating an interesting relationship between data and landscape may become increasingly available with the further development and application of sensor networks. The research summarised here aims to increase the understanding of the use of desktop based 3D virtual environments with a focus on the graphical representation of quantitative data through abstract symbols or graphics.

A mixed methods research approach is employed. Four different stages with different methodologies are combined to gain a holistic view regarding the goals of the study. The research stages are positioned along a 'bridge' from experimental 'in vitro' research to applied settings or 'in vivo' case studies driven by increasing context, data and task complexity. In the first stage, the effectiveness and efficiency of 2D bars in 3D virtual environments as compared to 2D displays was tested. Experiment participants identified the larger of pairs of bars and compared their lengths. The research stages IIa and IIb tested 2D bars in virtual environments with more complex data and tasks. In stage IIa participants answered complex tasks, such as pattern identification, in regard to several single value bars while in stage IIb a more open insight reporting approach was employed to let participants explore bar charts representing more complex data aggregations. The reported insights were analysed regarding their complexity, plausibility and the participants' confidence in them. In stage III a descriptive and explorative case study approach with three diverse cases including real world data sets and data experts was implemented to test and enhance the findings of the previous stages.

The results show that typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative data in virtual environments when represented as 2D bars on billboards. While the users are able to relate multivariate data represented in virtual environments to altitude and landform, the 3D environment does not especially support this. Only insignificant variation between 2D representations and 3D visualisations are found. However, the different data sets and tasks influence the results. The participants' answers are strongly guided by the tasks and some data sets are more successfully analysed in 3D, others in 2D. Generally, analysis of data in relation to altitude and landform is successful in either visualisation but participants do it less habitually than data analysis in relation to location and distribution. The data experts of stage III comment positively about the possibilities of the quantitative data visualisations in virtual environments. But the usefulness is dependent on visualisation completeness and on the data expert's previous usage of visualisations for either communication and/or data exploration purposes. Displaying up to four variables at once is identified as maximum of acceptable data graphics complexity. Additionally, more interaction, such as switching on and off the reference frames of the bar charts, is requested. Navigation is imperative for data analysis in virtual environments.

Methodologically bridging between experimental 'in vitro' and case study based 'in vivo' research methods is appropriate as the results of each stage can inform the design of the following stages. Additionally, the outcomes of later stages lead to re-evaluations or different interpretation of earlier results as for the aspects of bar chart complexity, occlusion and use of reference frames. Thus, in combining different methods, particular strengths such as exclusion or inclusion of context can be added together and potential weaknesses, such as small numbers of data experts, overcome. A holistic understanding of the visualisation technique is gained but it is nevertheless possible to attend to details. The case studies indicate that it is difficult to capture the use of visualisation in real world settings as the kinds of data sets made available are likely to be well known, as they were in this study. Nevertheless, the results of stage III allow evaluating earlier findings in a more applied setting and explore further issues. For example, the data experts commented on improvements and further applications for the visualisations. This may serve as input to the design process of future visualisation prototypes.

1. Introduction

Virtual Globes, such as Google Earth, are popular providers of desktop-based 3D virtual environments. Virtual environments are defined as a visualisation of part of the whole world based on a virtual globe technology. The development of the virtual globes of today's providers builds upon a history of visionaries and early technologies most of which no longer exist (1.1). 3D virtual environments often consist of a digital elevation model and high-resolution ortho imagery. Based on virtual globe technology a global reference system is provided making them suitable for the visualisation of transnational as well as local geodata sets. This research is based at the confluence of the existence of virtual environments, the availability of vast amounts of geodata (e.g. collected by geosensor networks) and the potential of geovisualization approaches which allow understanding data and making sense of them (1.2). Geosensor networks produce data sets which may be usefully analysed visually in relation to the landscape they were collected in. Additionally, easy data integration into virtual globe technologies has produced a number of data visualisations within virtual environments which are rarely evaluated for their appropriateness. 3D visualisations are very popular but they should not be used just because it is possible (1.2.1.4). Existing 3D visualisations can be grouped into three categories depending on their use of the x-, y- and z-axis of space (1.2.1). This research focusses on the third category where more or less realistic representations of the real world environment are enhanced with additional data displays. Thus, the x, y and z coordinates are used to show real world dimensions and additionally also data values (1.2.1.3).

Many research studies have evaluated desktop-based and immersive 3D visualisations by themselves or by comparing them to 2D representations (1.2.2). However, the studies often focus on the interface or navigation (e.g how useable not how useful), are very specific for an application area and/or evaluate specific tasks and data sets. A number of different approaches to display additional data within virtual environments are proposed but rarely formally evaluated. Generally, the reported research does not allow concluding that either 2D or 3D visualisations perform better as the results vary depending on tasks, information displayed and general display characteristics. In addition to the analysed research studies, a number of research agendas have outlined the potential of virtual environments for geovisualization and identified research challenges concerned with new technologies and new types of representations (1.2.3).

While data displays within virtual environments are popular, we know little about the appropriateness of such visualisations (1.3). Additionally, an ever increasing amount of geodata, for example collected through geosensor networks, is available. Such data sets may gain from visual analysis in relation to the landscape they were collected in. Evaluating the appropriateness of displays of quantitative data within virtual environments can thus provide the validation of an already used visualisation technique and, additionally, explore the possibilities for future visual analysis of data in relation to the landscape and especially to altitude and landform. The natural impression of landform in virtual environments is hypothesised to help its analysis, especially as traditional encodings of altitude and landform in 2D maps are difficult to interpret for some users. Current 3D geovisualization approaches are often technology driven and a holistic evaluation of the appropriateness of an existing visualisation technique seems most sensible. As a single research method is too limited, a combination of methods and different research stages which are designed driven by specific visualisation and application characteristics is required. The goals of this research are to increase understanding of the use of desktop-based virtual environments with a focus on quantitative data displays. In addition, it is aimed to relate experimental studies and case studies in specific application areas by evaluating this visualisation type for analysing quantitative data in a range of experimental and applied settings (1.4). Complementing these general aims a number of research questions and hypotheses were formulated (1.4). A summary of the main contributions of this research (1.5) ends this chapter.

Chapter 2. Methods >

1.1. Virtual Globes - the past and the present

Globally visible effects from events such as the volcanic explosion of Krakatao in 1883 stimulated an early global awareness (Dörries 2005). In the early 20th Century the vision of 'Spaceship Earth' showed up; a global view of the earth as a spaceship with a finite amount of resources. The term was mainly shaped by Richard Buckminster Fuller (1895-1983), an American architect, designer and futurologist, who published the book "Operating manual for spaceship earth" in 1969 (Fuller 1969). The term 'Digital Earth' has been popular since the speech "The Digital Earth: Understanding our planet in the 21st Century" by the former US vice president Al Gore on 31 January 1998 (Gore 1998). The main issues he mentioned were the need for a 'Digital Earth' which should be a three-dimensional representation of the planet earth, multi-dimensional and able to include the ever increasing amount of geo-referenced data which is collected throughout the world to make sense of it. This virtual representation of the earth should be connected with digital knowledge archives from all over the world and allow a better description and understanding of the system earth and also human activities (Gore 1998).

Technically, the development of computers and especially the increasing computing powers soon enabled visualisations in three dimensions and the invention of the World Wide Web and faster Internet connections support their distribution. VRML (Virtual Reality Modelling Language) and its successor X3D (Extensible 3D, W3D 2010) were early formats which allowed the distribution and display of smaller 3D data sets based on a local area flat earth model over the Internet. The early development of digital virtual globes was shaped by different European and American research institutes and companies such as Viewtec, GEONOVA, Keyhole, 3D GEO or GeoTango. All of these are either no longer existent or have been bought out by today's providers of virtual globes such as Google Earth (Google 2010, bought Keyhole), Microsoft Bing Maps 3D (Microsoft 2010*a*, bought GeoTango), Autodesk (Autodesk 2009, bought 3D GEO), NASA World Wind (NASA 2008), ArcGlobe (ESRI 2003), Virtual Explorer (Leica 2005) or rather ERDAS TITAN (ERDAS 2010).

Ten years after Al Gore's speech (Gore 1998) a number of scientists argue that the vision of the digital earth painted back then needs re-evaluation and offer their views as input for discussion (Craglia, Goodchild, Annoni, Camara, Gould, Kuhn, Mark, Masser, Maguire, Liang & Parsons 2008). They name four key developments over the last 10 years which are important for the definition of the next-generation digital earth. These four are the developments in spatial data infrastructures (SDIs) for the organisation of information, the different geobrowsers available for organising the information spatially, the developments in the geosensor technologies which allow 'sensing the World' and, last but not least, the innovations in supporting technologies such as faster hardware and internet connections. Their vision of the next-generation digital earth is based on the fundamental questions of our time - how does the earth's environment change and what are the consequences for human civilisation (Craglia et al. 2008). Based on this they propose that the next generation digital earth should not be one but rather multiple connected systems for different audiences and problems. It should offer various search possibilities through space and time and enable access to data and services while being engaging, interactive and exploratory and visualise not only the real world but also abstract concepts and data types.

Regarding the naming, for example 'digital earth', 'virtual globes', etc., Harvey (2009) offers a commentary on the different terms in use. He concludes that 'digital earth' is more suitable for an integrative discussion, while 'virtual globe' may rather be used for application software environments. An EuroSDR survey has shown that the terms 'virtual globe' or 'digital globe' are preferred by participants (Nebiker, Gülch & Bleisch 2010). In the research reported here the terms 'virtual environment' or '3D virtual environment' are used. Virtual environments are representations of smaller sections of the earth which are relevant for the application and tasks at hand and not a representation of the whole globe. However, technically they are based on virtual globe technologies able to represent the whole globe. This is especially important as cross-border data sets need a global spatial reference and, additionally, the findings of data analysis may need integration with other data sets making the use of local reference systems for the virtual environments impractical.

1.2. Motivation and background

The research presented here is located at the confluence of the availability of virtual globe technologies providing virtual environments, the availability of vast amounts of geodata (e.g. collected by geosensor networks) and the potential of geovisualization approaches which allow understanding data and making sense of them. Each of these aspects will be discussed briefly in the following sections.

1.2.0.1. Geodata availability

Today, vast amounts of data and information are available. Estimates suggest that 80% of all digital data comprise direct or indirect geospatial referencing, for instance, geographic coordinates, addresses, postal codes, etc. (e.g. MacEachren & Kraak 2001, VE 2007). The spatial reference enables integration of these data no matter what the sources are. Sensor Webs are a group of distributed sensors which are interconnected (figure 1.1) and share the data they collect through standardised interfaces (GeoICT 2007). Implementations of Sensor Webs and similar technologies already do, but certainly will in the future, generate vast amounts of data about our environments (Tao & Liang 2009, Nebiker, Christen, Eugster, Flückiger & Stierli 2007, Botts et al. 2006, Morville 2005). Gross (1999) predicted some years ago that "In the next century, planet Earth will don an electronic skin [. . .] will probe and monitor cities and endangered species, the atmosphere, our ships, highways and fleets of trucks, our conversations, our bodies – even our dreams." Even though this might not become true in every 'last' detail we are nevertheless challenged to find ways to explore and analyse all the collected data and transform it into information and later into knowledge (Thomas & Cook 2005, MacEachren & Kraak 2001).

1.2.0.2. Geovisualization

Geovisualization is concerned with visualisations for the exploration, analysis and communication of spatially related data and information also called geo data. MacEachren & Kraak (2001, p. 3) define "Geovisualization integrates approaches from visualisation in scientific computing (ViSC), cartography, image analysis, information visualization, exploratory data analysis (EDA), and geographic information systems (GISystems) to provide theory, methods, and tools for visual exploration, analysis, synthesis, and presentation of geospatial data (any data having geospatial referencing)". Thus,

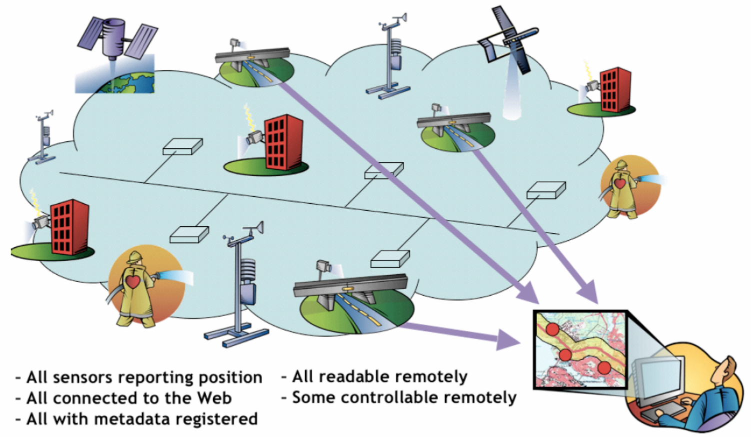


Figure 1.1.: Sensor Web concept (Botts et al. 2006, p. 4)

the display and analysis of data in virtual environments is also a geovisualization approach. In this research the focus lies on geovisualization for data exploration and analysis, and not for data communication.

1.2.0.3. Virtual environments

Virtual environments are computer-based interactive 3D representations of real (e.g. figure 1.2) or artificial landscapes and/or objects that invoke a sense of realism (Slocum, Blok, Jiang, Koussoulakou, Montello, Fuhrmann & Hedley 2001). MacEachren, Kraak & Verbree (1999) define "a GeoVE as any virtual environment [both desktop and non-desktop/immersive] used to represent geospatial information (either measured or simulated). Thus GeoVEs are virtual environments that represent characteristics of the world (or possible worlds) at scales from the experiential (e.g. a neighbourhood) to the global." Additionally, MacEachren, Edsall, Haug, Baxter, Otto, Masters, Fuhrmann & Qian (1999*a*, p. 36) state that in GeoVEs it is possible to "depict more than the visible characteristics of geographic environments [. . .] to produce geospatial virtual 'super environments' in which users can not only see what would be visible in the real world, but also experience the normally invisible". Popular providers of 3D virtual environments are either virtual globe technologies such as Google Earth (Google 2010) or i3D (Nebiker & Christen 2010), which employ global coordinate systems (e.g. the World Geodetic System 1984 WGS84, NGA 2010), or tools and languages for the representation of parts of the earth (e.g. Cinema 4D (MAXON 2011) or X3D (W3D 2010)), which mainly use local coordinate systems. In this research the terms 'virtual environment' or '3D virtual environment' are used to mean a desktop-based (viewed on the 2D computer screen) 3D virtual environment displaying a 3D landscape (mainly consisting of a digital terrain model and draped high resolution ortho imagery) independent of the underlying technology. Desktop-based 3D displays trick us into seeing the virtual environment in 3D by using monocular depth cues (Ware 2004).

1.2.1. Categorisation and examples of 3D representations and applications

Manually created 3D representations of the real world have existed for a long time, for example, the panoramas of Berann (Troyer 2007) or historic city maps (e.g. FOXNews.com 2010, de Boer 2008). A short overview of the history of using the third dimension in maps can also be found in (Kraak

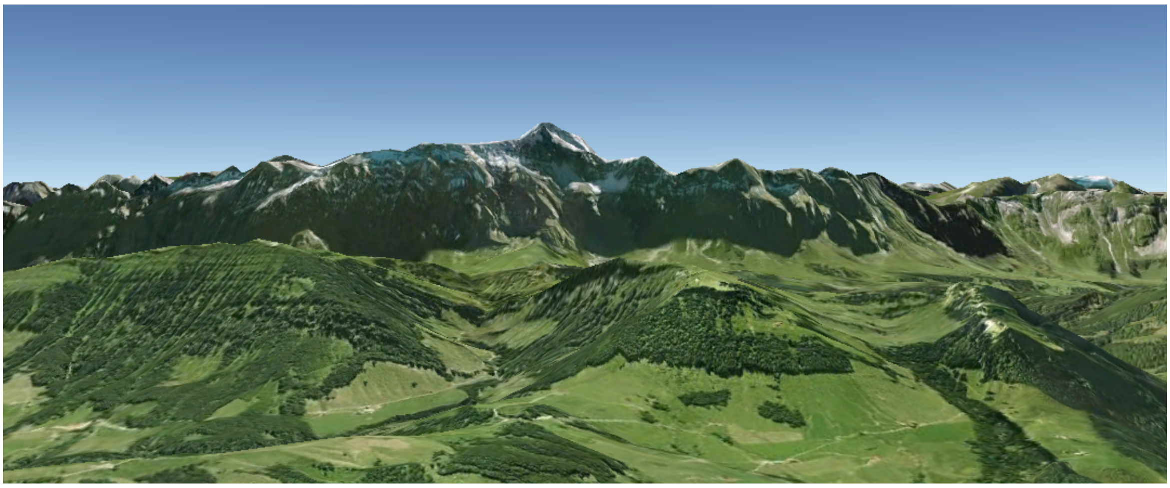


Figure 1.2.: Virtual environment showing a real landscape - view of the Säntis mountain in eastern Switzerland in Google Earth (Google 2010)

1988). The research reported here concentrates on digital 3D representations. For differentiation 3D representations can be categorised. Elmqvist & Tudoreanu (2007) distinguish between two reasons for creating 3D 'virtual worlds': 1) replicating the real world and 2) using the 3D as a canvas for abstract information. Their categorisation leaves out the option that sometimes abstract information and real world displays are combined. The following examples of 3D representations are grouped in three categories even though the boundaries between the categories are not clear cut. The three categories based on Elmqvist & Tudoreanu's (2007) distinction are: 3D representations with a ...

1) ...focus on data - scientific 3D visualisation for data only

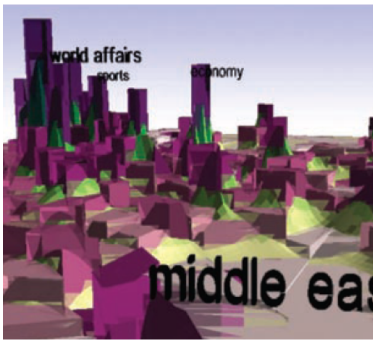
2) ...focus on depicting the real world (environment and objects)

3) ...combination of 1) and 2), displaying data or abstract information within virtual environments

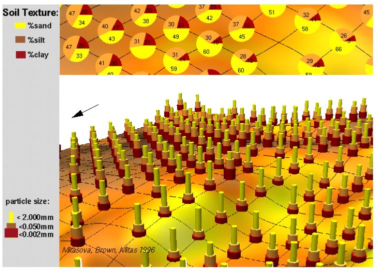
The following three sections explain and give examples of 3D representations in each of these three categories.

1.2.1.1. Examples for 1) - focus on data

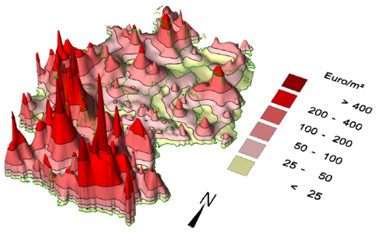
In this category the 3D representations display mainly data and/or objects of the real world (e.g. parts of the human body for medicinal visualisations) most often in a local coordinate system. The x, y, and z coordinates of the display are mainly used to show data values and, where needed, some context information. Most scientific visualisations belong to this category. Examples can be found in figure 1.3. Figure 1.3a shows a data only 3D display that employs spatial metaphors for data communication, called spatialization. The next example (figure 1.3b) shows soil texture data. An interpolated 3D data surface is enhanced with point data which is also displayed in 3D (explaining the soil texture at each point). This representation is combined with a traditional 2D display with pie charts within the same graphic. Figure 1.3c shows land prices in Germany. While the map of Germany is used as a reference the display does not aim to represent Germany but rather focuses on the representation of the land price data and the map as a background helps the georeferencing or localisation of the findings.



(a) Spatialization (Skupin & Fabrikant 2003)

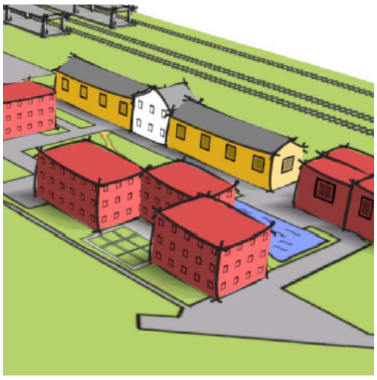


(b) Soil texture (Mitas et al. 1997)

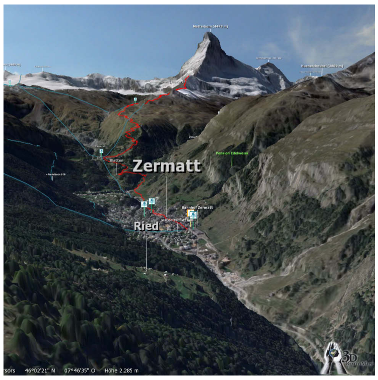


(c) Land prices (Rase 2003)

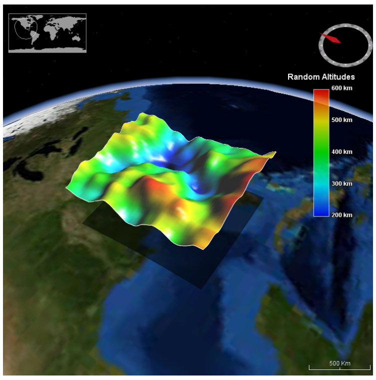
(d) Virtual Berlin 3D (Berlin 2010)



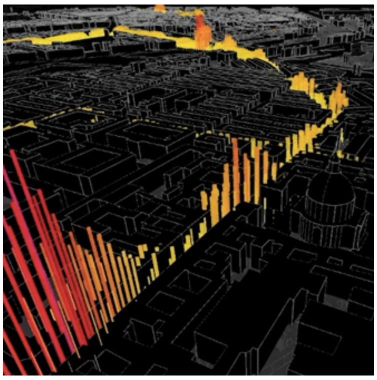
(e) Cartoon city model (Döllner & Walther 2003)



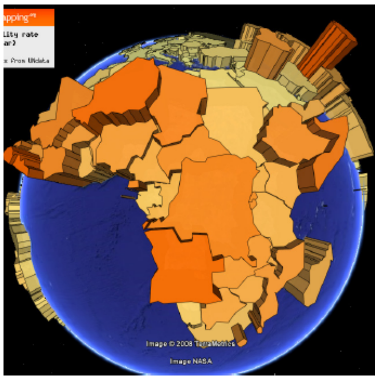
(f) 3D Reality Maps, Zermatt Demo (RealityMaps 2010)



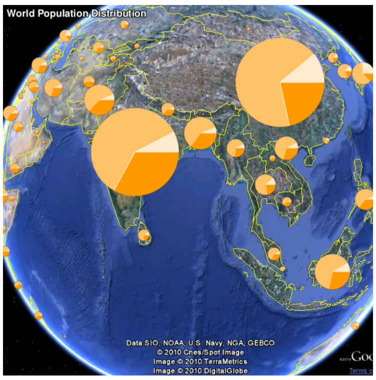
(g) Analytical surface in NASA World Wind (NASA 2010)



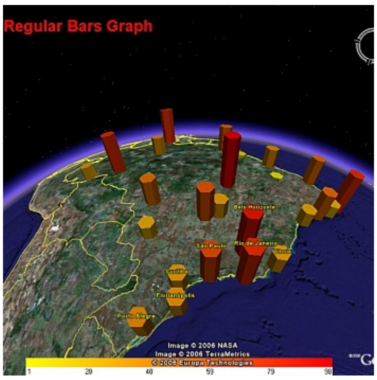
(h) Copenhagen Wheel sensor data (Ratti et al. 2010)



(i) Thematic mapping engine: prism map (Sandvik 2010)



(j) Thematic mapping engine: pie charts (Sandvik 2010)



(k) GE-graph: 3D map (Sgrillo 2010)

(l) Insurance data, in transition from one display state to another (Slingsby, Dykes, Wood, Foote & Blom 2008)

Figure 1.3.: Examples of 3D representations

1.2.1.2. Examples for 2) - focus on depicting the real world

This category comprises of virtual environments representing the real world and/or its objects in a realistic or abstract/generalised way. The x, y, and z coordinates of these displays are mainly used to show the real world dimensions easting, northing and elevation or the dimensions, including height, of buildings or other objects. MacEachren, Edsall, Haug, Baxter, Otto, Masters, Fuhrmann & Qian (1999*a*) call these virtual environments ”spatially iconic GeoVEs”. Digital city models such as in figure 1.3d and virtual globes such as Google Earth are typical examples of this category of 3D representations. There is much research going on focussing on the detailed construction and realistic visualisation of city models (cf. Nebiker, Bleisch & Christen 2010, for an overview). Other approaches aim to visualise the real world city models in a more abstract way (e.g. Döllner & Walther 2003, in figure 1.3e). Virtual environments aiming to represent a real environment typically consist of a digital elevation or surface model with some sort of drape. For virtual environments which are designed to look realistic (e.g. figure 1.2) the drape typically consists of high resolution ortho imagery (Lange 1999) but satellite imagery or maps could also be used. Figure 1.3f shows a tourism example for vacation planning (Siegert 2010). A digital surface model (instead of a digital elevation model, for a more realistic representation of forests and large rocks on the slopes) is overlaid with high resolution ortho imagery and enhanced with some additional information such as labels for place names, tourist infrastructure and overlaid hiking routes. Users of virtual globes such as Google Earth or MS Bing Maps 3D (Microsoft 2010*a*), are probably most used to the type of 3D representations belonging to this second category.

1.2.1.3. Examples for 3) - combination of data and real world representation

The third category enhances the more or less realistic representation of the real world environment with additional data displays. Here the x, y, and z coordinates are used to show real world dimensions and additionally data values. Examples for this category are the analytic data surface overlaid over NASA World Wind (figure 1.3g), the 3D representation of the data collected by the Copenhagen Wheel (Ratti et al. 2010) within the virtual city of Copenhagen (figure 1.3h) or the 3D visualisation of global statistics in Google Earth (figure 1.3i). Beside the ’well-known’ virtual globes such as Google Earth or NASA World Wind there are also smaller initiatives which allow a combined visualisation of the environment or landscape with additional data such as the i3D Virtual Globe technology (Nebiker & Christen 2010) or the SPI Operational Environmental Emergency Response Tool (SPI 2007). But this is far from being an exhaustive list. In general, the type of 3D visualisations shown in these examples are the ones this research focusses on - displaying additional data (e.g. measured values from a sensor network) within a realistic looking desktop-based 3D virtual environment depicting part of the real world.

1.2.1.4. Discussion

The borders between the categories defined above are not fixed, neither is there a convention for naming them. For example, Polys & Bowman (2004) name their mainly scientific 3D visualisations enhanced with labels and some other additional information ”information rich virtual environments” or Bodum (2005) offers a categorisation of virtual environments in geovisualization based on the degree of realism/immersion and temporal characteristics. Defining the categories in section 1.2.1

for this research helps to distinguish between different visualisation types and define the focus of this research.

A similar categorisation of "spatial iconicity" for virtual environments is presented by MacEachren, Edsall, Haug, Baxter, Otto, Masters, Fuhrmann & Qian (1999*a*, p. 36). Their three generic categories 'abstract', 'iconic', and 'semi-iconic' approximately match the categorisations presented above in this order (section 1.2.1). However, their definition of 'semi-iconic' virtual environments maps an abstract data value to one of the geographic dimensions (e.g. as done in space-time cubes visualising time-geography, introduced by Hägerstrand (1970), more details in Parkes & Thrift (1980), or with data surfaces) and does not allow for double use of one or several dimensions for depicting the real world and additionally abstract data values as in category 3) above (section 1.2.1.3).

Creating visualisations belonging to the third category, which combine the visualisations of the real world with data displays, is helped by a number of tools. Especially for the virtual globe Google Earth there are tools available which make it easy to integrate data into the virtual environment. GE-Graph (Sgrillo 2010) allows easy generation of diverse graphs and data displays for Google Earth (figure 1.3k). The Thematic Mapping Engine developed by Sandvik (2010) for his MSc thesis allows visualising global statistics in 2D or 3D on Google Earth (figure 1.3j). Many other ways such as using scripting languages to access the Google Earth API for creating 3D representations are also available. Section 2.3.2 of this report describes an XML based process for the integration of data displays in Google Earth. A further example (Slingsby, Dykes, Wood, Foote & Blom 2008, figure 1.3l) uses Google Earth to show French insurance data. Such a data set does not especially call for a 3D background. So why was it implemented in a virtual globe technology? Google Earth offers easy data integration with the KML language and the freely available virtual globe viewer has the potential to make the data representation available to a large audience. Easy data integration and free viewing is something that is often missing from many tools used for the (expert) display and analysis of geodata such as commercial GIS software. Slingsby, Dykes & Wood (2008) also offer a tutorial explaining how to easily integrate data into Google Earth.

These two aspects, easy data integration and free availability, have the potential to make a tool widely used and geobrowsers have become a de facto standard for visualising spatial information on the desktop (Wood, Dykes, Slingsby & Clarke 2007). The Wallpaper (2007) has even awarded Google Earth the title "most life-enhancing item" and Walsh (2009), besides giving 35 graphic examples, states that "3D is the new 2D." The popularity of 3D displays is also shown by Bartoschek & Schönig (2008) who did a study on the streets of Münster, Westfalen where they found out that 65% of the participants are familiar with virtual globes such as Google Earth. 3D displays are popular in the scientific/research domain too. In 2006, just after the introduction of Google Earth, Butler (2006, p. 776) asked "Life happens in three dimensions, so why doesn't science?" Analysing the front cover topics of the KN (Kartographische Nachrichten - the German journal for cartographic research) shows that, depending on the definition of 3D, 6-9 of 12 cover pages of the journal Kartographische Nachrichten from mid 2008 to mid 2010 featured some sort of 3D representation (figure 1.4). Also the report on the EuroSDR Project on Virtual Globes (Nebiker, Gülch & Bleisch 2010) shows a strong current use of virtual globes for viewing standard and local/personal contents and foresees this and more geospatial collaboration uses for the future.

It is concluded that 3D representations are very popular and often used especially because most viewers are freely available and data can easily be integrated. The definition of three categories helps differentiate between varied types of 3D representations. However, even though 3D is very popular it is not well known if and when 3D has a distinct advantage over other types of visualisations, especially 2D visualisations. Shepherd (2008, p. 200) remarks that sometimes a "3D for 3D's sake" tendency is apparent. The review in the following section 1.2.2 summarises the results of a number of studies evaluating various aspects of 3D visualisations. As data visualisations within virtual environments (the third of the above defined categories, section 1.2.1.3) have rarely been implemented and evaluated so far, many studies of 3D visualisations of the other two categories are included in the review especially if the researched aspects are relevant to the evaluation of 3D representations belonging to the third category.

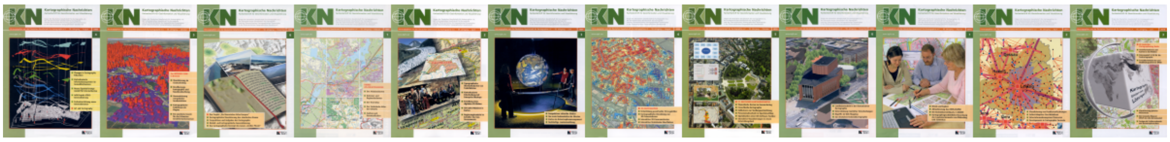


Figure 1.4.: Front covers of the journal Kartographische Nachrichten from mid 2008 to mid 2010

1.2.2. Evaluations of 3D representations and applications

One of the earlier studies on 3D interfaces was Robertson, Czerwinski, Larson, Robbins, Thiel & van Dantzich's (1998) research on the document management system data mountain which showed that the 3D data mountain has significant advantages over MS Explorer. The idea of the data mountain was further tested by Cockburn & McKenzie (2001) who found that there is no significant difference between the 2D and 3D interface and that generally the document retrieval time increases with the increase of the number of documents. However, the same study showed a significant preference of the 3D interface based on Robertson et al.'s (1998) data mountain by the participants. Questioning these results obtained with static 3D interfaces, Zhu & Chen (2005) did a similar study on retrieving knowledge from a repository employing an interactive 3D interface. They concluded similarly that the 3D interface is at least as effective as the 2D interface. However, more interaction is needed in 3D and there is the problem of hidden objects in the 3D interface (Zhu & Chen 2005). Sebrechts, Cugini, Vasilakis, Miller & Laskowski (1999) evaluated the visualisation of search results in 2D, 3D and text form finding that 3D comes at high costs such as high mental load for navigation and interaction. They also found that the mental load decreased with increasing familiarity with the 3D visualisation. Thus they suggest not to evaluate 3D visualisations using short term studies with novice users. Another finding of Sebrechts et al.'s (1999) study showed the dimensionality of the visualisation seemed to matter less than the tasks set and the available features. Cockburn & McKenzie (2004) later revisited their findings asking if spatial memory is better supported by 2D or 3D. They list several studies which have concluded that 3D supports spatial memory. An experiment controlling for previously uncontrolled factors shows that spatial memory is important but probably not aided by 3D depth cues (Cockburn 2004). Wickens, Olmos, Chudy & Davenport (1997) summarise earlier comparisons of 2D and 3D displays in aviation and state that the benefits and costs of 3D displays are complex issues which depend on the tasks, the displayed information and the rendering. Similarly Keehner, Hegarty, Cohen, Khooshabeh & Montello (2008) found that the visibility of task relevant information is crucial for performance, while the participant's active control of the visualisation does not enhance task performance.

Dawood & Sikka (2007) evaluated the effectiveness of 4D planning (3D models plus schedule) for communication compared to 2D CAD models finding that the 4D participants where faster and better able to communicate. Similarly positive results for 3D are reported by Irani & Ware (2003) who compared the interpretation of 3D diagrams (using shaded primitives) to 2D UML diagrams. They found that in 3D the substructures of the diagram could be identified much faster and recalled more reliably than in 2D. Shneiderman's (2003) article titled ”Why Not Make Interfaces Better than 3D Reality?” critically reviews the controversy about 2D versus 3D interfaces giving some of the first guidelines for 3D designers. He concludes saying that ”three-dimensional environments are greatly appreciated by some users and are helpful for some tasks. [. . .] Success will come to designers who provide compelling content, relevant features, appropriate entertainment, and novel social structures. Then by studying user performance and measuring satisfaction, they can polish their designs and refine guidelines for others to follow.” (Shneiderman 2003, p. 15)

But not only 3D interfaces to data are studied and evaluated there are also a number of evaluations of 2D versus 3D displays. Most often they are done with basic tasks and in specially designed and tightly controlled experiments, rarely leaving room for context information or tacit knowledge, as is abundant in real world settings and applications, to influence the outcomes.

Risden, Czerwinski, Munzner & Cook (2000) compare 2D and 3D visualisations of web content finding that the results in 3D are faster and with the same quality as in 2D. Asking for further research they would like to know how 2D and 3D displays could be optimally combined. Also St. John, Cowen, Smallman & Oonk (2001) see a great potential for combining 2D and 3D views especially for more complex tasks involving shape understanding and relative positioning. In their study they evaluate desktop-based 2D and 3D displays regarding shape understanding (identification and mental rotation) and precise judgement of relative position (locating shadows and determining directions and distances between objects), two task areas which they think important for applications such as air traffic or military control and command and potentially for any other display of 3D information. Additionally, they review 16 studies from the 1990s which empirically test 2D and 3D displays concluding either that 2D or 3D are better or that both display types perform equally. For their own study St. John et al. (2001) conclude that the integration of dimensions in 3D displays facilitates shape understanding but on the other hand the distortions inherent in 3D displays hamper judging relative positions. Tory, Atkins, Kirkpatrick, Nicolaou & Yang's (2004) eye gaze analysis on how best to arrange 2D and 3D displays might be helpful for the optimal use of combined 2D and 3D displays. They found that the 3D display was used more often than 2D and recommend having a 3D overview in the middle of the display.

Schnabel (2003) evaluates the perception of a 3D cube and a 3D maze in a virtual environment compared to conventional architectural 2D depictions and finds that 3D enhances the understanding of spatial issues and enables a better comprehension of complex volumes and their spatial relationship. More differentiated results are found by Tory, Kirkpatrick, Atkins & Möller (2006) who conducted a series of experiments comparing 2D, 3D and combined 2D/3D displays for relative positioning and approximate navigation tasks. They found that 3D displays are only effective for precise navigation and positioning when good viewing angles or measurement tools are available. For precise tasks in other situations combined 2D/3D displays showed better performance and facilitate higher confidence. Thus, several factors such as task characteristics, orientation cues and occlusion influenced the performance of the different displays (Tory et al. 2006). Also Forsberg, Chen & Laidlaw (2009) found that task performance varied together with the provided depth cues, occlusion and the clarity of the visualised information. They tested four different 3D vector field visualisation methods with five simple but representative tasks. In a study by Swienty, Jahnke, Kumke & Reppermund (2008) 2D proved to be superior to 3D for the visual scanning of geographic information in visualisations based on the attention-guiding principles, which is focussed on the location of information but not on the information itself. For 3D they propose a model where relevant information should be highlighted and state that further research in 3D is needed.

Wang, Bowman, Krum, Coelho, Smith-Jackson, Bailey, Peck, Anand, Kennedy & Abdrazakov (2008) compare video placement within 2D and 3D contexts regarding path reconstruction tasks. They find that the 3D model enabled realtime strategies and led to faster performance. Also participants who were not familiar with the environment achieved a similar task performance as participants working in the displayed environment. On the other hand they remark that 2D is simpler and easier to learn (Wang et al. 2008). Comparing the 3D virtual environment to the real world for similar tasks (estimating walking path lengths) Jansen-Osmann & Berendt (2002) conclude that desktop-based virtual environments are a valid and economic research tool which can replace such research in the real world. They base their conclusion on the findings that the participants achieved the same results in the virtual environments as were reported from the same experiment in physical spaces. Another study (Lorenz, Thierbach, Kolbe & Baur 2010) evaluates the use of 2D maps and 3D visualisations for indoor navigation and proposes conclusions for the design of indoor navigation maps.

In a number of application areas such as geology, air traffic control, medicine or simulations the use of 3D visualisations is quite established. Jones, McCaffrey, Clegg, Wilson, Holliman, Holdsworth, Imber & Waggott (2007) report on two case studies visualising multi-scale geological models in 3D and concludes that graphical user interfaces which are based on a virtual world metaphor provide the best user interactivity. Comparing the planning of a liver surgery in 2D and desktop-based 3D, the 3D system better supports the surgeons (Reitinger, Bornik, Beichel & Schmalstieg 2006). Also Basdogan, Sedef, Harders & Wesarg (2007) state that VR-based simulators for training in minimally invasive surgery are a promising alternative to traditional training based on their discussion and analysis of 31 commercial simulation systems. For air traffic control or aviation displays in general, many studies regarding the testing and evaluation of 3D visualisations can be found in aviation journals (e.g. Haskell & Wickens 1993) but the findings are not always clearly for or against 3D. A study by Mejdal, McCauley & Beringer (2001) recommends the use of 3D for traffic displays and navigation displays but not for weather displays. They recommend this because 3D is more intuitive and natural but not without drawbacks such as difficulties in depth judgments, line-of-sight problems or occlusion (Mejdal et al. 2001). Also Smallman, St. John, Oonk & Cowen (2001) evaluate desktop-based 3D displays and the information availability in such displays for air traffic and piloting applications. Based on a number of studies concluding that the rapid comprehension of information about the third dimension is better in 3D displays than in 2D displays they designed their own study with stricter controlling and found that 2D is faster and that the information availability (not only realism) in 3D displays and the type of coding has a great influence on task times. Another area where 3D visualisations are established is the gaming industry. Many of the 3D geovisualization applications even profit from developments for the gaming industry such as faster hardware, advances in computer graphics or interaction devices (Mine 2003, Weber, Jenny, Wanner, Cron, Marty & Hurni 2010) even though the requirements on geometric and visual accuracy and precision in gaming and geovisualization normally differ greatly.

A number of studies proposing new applications, methods and techniques for 3D representations are also interesting. Unfortunately, some of them have not been formally evaluated yet. For example, Kreuseler (2000) describes in detail different techniques how additional data can be displayed in virtual environments in their developed system. One of his examples includes the displays of bar charts as are evaluated in this study. However, Kreuseler's (2000) system is neither formally nor informally evaluated even though he admits that the visualisation of data in virtual environments is complicated and not widely studied. Other un-evaluated 3D visualisation applications include Pelechano & Badler's (2006) 3D visualisation for showing the results of crowd simulations during building evacuation, Wang's (2005) description of the challenges and benefits of integrating GIS, simulation models and 3D visualisations into one prototype system for traffic impact analysis or Hiebel, Hanke & Hayek's (2010) 3D representation of project data. Proposed new algorithms and implementations of view deformations and other projections in 3D virtual environments (Yang, Chen & Beheshti 2005) might also benefit from evaluation as well as Qu, Wang, Cui, Wu & Chan's (2009) implementation of a new focus & context route zooming and information overlay technique for 3D visualisations. On the other hand, Tiede & Lang (2010) collect user feedback for their so-called analytical 3D views in virtual globes and Kettunen, Sarjakoski, Sarjakoski & Oksanen (2010) at least plan to evaluate their proposed oblique parallel projection for cartographic 3D representations. The widespread use of 3D representations also leads to 3D specific research and evaluations. For example, Elmqvist & Tsigas (2008) review the occlusion management techniques developed over recent years and provide a comprehensive taxonomy of occlusion management techniques. Maas, Jobst & Döllner (2007) observing the problems with the extensive use of labelling in virtual environments which may destroy important depth cues and thus potentially impede human perception, discuss and implement a new labelling algorithm supporting depth cues and propose the evaluation of it.

Another large research area, which shall not be discussed in detail here but is mentioned, are the immersive 3D virtual environments such as the CAVE or stereoscopic viewing. Schratt & Riedl (2005) describe different 3D display technologies. Generally, the main issues for immersive systems are navigation and interaction. Elmqvist & Tudoreanu (2007) compared an immersive CAVE environment with desktop-based 3D virtual environments finding that user performance (accuracy) is not influenced by the display technique. However, navigation behaviour and time spent strongly depends on the display type. They conclude that the techniques have complementary properties (Elmqvist & Tudoreanu 2007). Also Polys, North, Bowman, Ray, Moldenhauer & Dandekar (2004) in their study about information rich virtual environments identified critical usability concerns for 3D immersive displays (CAVE). The Adviser prototype, an immersive cave environment for planetary geoscientists and geologists, was evaluated in five case studies (Forsberg, Prabhat, Haley, Bragdon, Levy, Fassett, Shean, Head, Milkovich & Duchaineau 2006). Based on the case studies they found that the understanding of the 3D terrains is clearer in 3D, that 3D helps better spatial judgements, that effective quantitative measurements can be made and that in 3D details can be seen which are overlooked or under appreciated in 2D (Forsberg et al. 2006). A study by Ware & Mitchell (2005) compared stereoscopic displays, 3D displays with kinetic depth cues and 2D displays for the visualisation of path graphs. Based on the analysis of similar studies they use a higher resolution for their displays and find that much larger graphs can be read in 3D (stereoscopic viewing performed even better than 3D displays with kinetic depth cues) than in 2D (Ware & Mitchell 2005). Johns (2003) concludes less in favour of immersive environments recommending non-immersive desktop-based 3D virtual environments for the teaching of spatial concepts.

Other evaluations of 3D virtual environments are concerned with the visualisation of the real world rather than with the visualisation of data and information in the third dimension. For example, Purves, Dowers & Mackaness (2002) evaluate a virtual reality setting assuming that the shape of the ground is the primary element of such a setting. They conclude that draping a map over the 3D landscape is only another representation of the map but they admit that it might be visually more appealing. Additionally, they added 3D objects, such as trees, finding that this added no new information but significantly degraded performance (Purves et al. 2002). Thus, they confirm the findings by Lange (1999) who concluded that drape is more important than 3D objects. In a similar study Wood, Pearson & Calder (2007) compared different 3D representations for wilderness navigation in Scotland. They found no significant differences for draped or undraped models. However, their participants preferred the draped models (Wood, Pearson & Calder 2007). Recently also Schobesberger & Patterson (2008) evaluated the effectiveness of 2D and 3D trail head maps for cartographic information communication and map preference by hikers. The results varied, thus they recommend making the decision for 2D or 3D maps on a case-by-case basis depending on the type of landscape and the intended audience. Another study by Bleisch & Dykes (2008) evaluated 3D virtual environments with some additional information, such as hiking routes, for the planning of hikes finding that overview tasks are better supported by the 3D visualisation than exact hike planning and route finding tasks.

Regarding the display of data in 3D, the evaluation of symbology and the resulting guidelines are helpful even though there are not so many studies done in this area of research. An early study by Kraak (1988) analyses different 3D maps mono- and stereoscopic. He concludes that the stereoscopic view leads to faster (especially for point symbols) but qualitatively similar results. Regarding symbology, he offers some guidelines as to what combinations of visual variables and depth cues are helpful or better avoided. He concludes that ”referring to cartography as a whole it can be said that the general cartographic theory can be applied [. . .]. In the cartographic communication process these maps, provided they are kept relatively simple [. . .], function at least as well as two-dimensional maps. For more complex maps further comparison between two- and three-dimensional maps will be necessary.” (Kraak 1988, p. 106) The work of Kraak (1988) is taken up by MacEachren (1995) in his book ”How Maps Work” and integrated into a comprehensive overview regarding the sensible use and combination of visual variables and depth cues for 3D maps. Newer books (e.g. Slocum, McMaster, Kessler & Howard 2005, Kimerling, Buckley, Muehrcke & Muehrcke 2009) recommend different types of mainly solid 3D symbols or visual variables for the use in thematic mapping. A study by Häberling (2003) set out to find design principles for topographical 3D maps. The outcomes of his expert reviews rank map like symbolisation higher than realistic representations. He also recommends a dynamic/interactive use of 3D maps (Häberling 2003).

As already mentioned above (e.g. St. John et al. 2001, Tory et al. 2006) 2D and 3D displays may be used in combination where each display type can offer its characteristics strength. They could be used beside each other or in sequence where Hollands & Ivanovic (2002) found that for military displays the performance with the 2D and 3D views is better when the displays undergo a continuous transition. Brooks & Whalley (2008) implement a multi-layer hybrid visualisation (the landscape and additional information) as combined 2D and 3D views. The transformation between the views in their study is continuous and under the control of the user. They claim that the combination of different displays takes advantage of the different strengths but they do not formally evaluate it (Brooks & Whalley 2008). Schafer & Bowman (2005) evaluated in a case study a prototype which integrates 2D and 3D views for spatial collaboration. They confirm that multiple representations of the same space are useful (Schafer & Bowman 2005). Chang, Wessel, Kosara, Sauda & Ribarsky (2007) found that 3D in combination with 2D displays supports gaining more understanding and deeper insights in the visualised urban relationships. And Jianu, Demiralp & Laidlaw’s (2009) study of combining 3D and 2D displays for the exploration of complex fiber tracts adds to this as they found the navigation within their complex displays easier when 2D and 3D are combined.

Characteristics, strengths and weaknesses of 3D visualisations are compiled in several reports. In addition to the advantages and disadvantages of 3D visualisations already mentioned above, Morrison & Purves (2002) give an overview of reasons why to use 2D and 3D representations of the landscape. Some of them are that 2D is familiar to many users and that is easy to navigate or that

3D is visually attractive and requires little interpretation of the form of the landscape. Medyckyj-Scott (1994, p. 207) state that 3D representation ”reduces the gulf of evaluation by being in some sense natural”. Similarly Wood, Kirschenbauer, Döllner, Lopes & Bodum (2005) remark that navigational and behavioural realism can be beneficial when virtual environments are used and Rase (2003) suspects that 3D views may be helpful as inexperienced map readers are not able to decode the encodings of height information on a 2D map, such as contour lines or hachures (Collier, Forrest & Pearson 2003). While Bleisch & Dykes (2008) found that virtual environments are more useful for overview than for detailed hike planning tasks, Nielsen (2007) attributes 3D environments with only a partial usefulness for overview tasks. She found that virtual environments are especially beneficial for supporting imagination and emotion or creating attraction. Shepherd (2008) reports an interesting and thorough review of three-dimensional 'geographical visualizations' by discussing reasons for and problems with 3D displays. The examples he shows are data visualisations within map-like flat earth 3D visualisations. This potentially also because he considers the multiple use of the z-axis of a 3D visualisations for the vertical space dimension and, additionally, one or even several data dimensions as problematic. However, the examples he provides in illustration of this problem include only 3D displays comprising multiple stacked data layers where the upper data layers would need visual subtraction of, for example, the terrain base layer (Shepherd 2008).

From the reported research findings above it is not possible to say generally that either 2D or 3D visualisations perform better. While some studies found evidence for better performance in 3D, others concluded that there is minimal difference between 2D and 3D or that 2D performs better. Often the results seem strongly dependent on the tasks, the information displayed and general display characteristics. The latter is especially obvious when comparing displays of older studies with newer ones. The advances in computer hardware and software have changed dramatically, for example, the interaction speed, the resolution and rendering quality and thus also the impression of 3D displays. It might even be the case that some older findings would need re-evaluation with newer displays. Evaluations of data displays within virtual environments are rare. Some implementations are reported but evaluation of them are often only proposed so far. A research area where some consistency in results is found is the combination of 2D and 3D displays. While 2D and 3D displays are combined in various ways it is generally concluded that concurrent multiple representations of the same space and/or data set is useful for gaining insight and/or for navigating the displays.

1.2.3. Review of research agendas

In addition to the various studies evaluating 3D displays (section 1.2.2), a number of research agendas have also outlined the potential of virtual environments for geovisualization and identified research challenges concerned with new technologies and new types of representations which directly relate to the research reported here. Gahegan, Wachowicz, Harrower & Rhyne (2001) note that for exploratory visualisation we should take advantage of newer developments such as virtual environments and research the effects of the visual environment on the knowledge discovery process. Geovisualization methods may facilitate science and decision-making in real world applications (Slocum et al. 2001). Similarly, MacEachren, Edsall, Haug, Baxter, Otto, Masters, Fuhrmann & Qian (1999*b*) propose research on applications that take advantage of the potential of virtual environments facilitating collaboration for decision-making. We need to investigate and use geospatial virtual ”super environments” to learn about possible advantages of them over 2D displays (MacEachren, Edsall, Haug, Baxter, Otto, Masters, Fuhrmann & Qian 1999*b*). Slocum et al. (2001) propose the analysis of "approaches to exploring geospatial data interactively in non-immersive desktop environments".

Further research challenges are the implementation and information display issues such as the content creation for virtual environments or the embedding of different types of abstract information in the 3D environment (Bowman, North, Chen, Polys, Pyla & Yilmaz 2003). Bowman et al. (2003) also mention the challenge of maintaining the legibility of embedded information of different types or the appropriate level of detail. Slocum et al. (2001) note the need for the integration of visible-tangible data about landscapes and non-visible and abstract data, the mix of realism and abstraction. They go on stating that appropriate mixes of e.g. cartographic, graphic and statistical approaches to understand geodata and the variation of the mix with applications need to be determined. Gahegan et al. (2001) ask for the types of visual encoding that work best for particular tasks, applications and data and MacEachren, Edsall, Haug, Baxter, Otto, Masters, Fuhrmann & Qian (1999*b*) state the need for an appropriate balance between realism and abstraction in different application domains.

These are only a sample of research challenges identified in recent years. Further challenges and research questions relating to these topics can be found in MacEachren & Kraak (2001), Polys & Bowman (2004), Johnson (2004), Johnson, Moorhead, Munzner, Pfister, Rheingans & Yoo (2006), Thomas & Cook (2005), Cartwright, Crampton, Gartner, Miller, Mitchell, Siekierska & Wood (2001), Fairbairn, Andrienko, Andrienko, Buziek & Dykes (2001), MacEachren, Kraak & Verbree (1999), Jobst & Germanchis (2007), Batty & Smith (2002) or Virrantaus, Fairbairn & Kraak (2009).

Kraak (2006) notes that most of the research challenges are still unmet. Or as Slocum et al. (2001, p. 14) put it "the most sophisticated technology will be of little use if people cannot utilise it effectively". Craglia et al. (2008), in their recent evaluation of the digital earth vision, propose among other topics that research on the visualisation of abstract concepts and data types in space is an important issue for the next-generation digital earth to become true.

1.3. Problem statement

Most data has a geospatial component, especially data collected with the help of geosensor networks which are increasingly common (section 1.2). Geospatial data is usually represented graphically using maps or it can be analysed through geovisualization techniques that support spatial sense-making. The geospatial component of data is often reduced to two-dimensional location as part of this process. There are some advantages in this simplification in terms of data processing, perception and cognitive load (MacEachren 1995). However, elevation may also be important in geographic data analysis, especially for data sets collected in mountainous environments and in analytical tasks where altitude, altitude differences and landform could be considered to get deeper insights into those data sets. Thus, a visual combination of data displays with the virtual environment depicting the real area where the data set was collected (section 1.2.1 category 3 for potential examples of such displays) may be helpful for analytical tasks considering data in relation to altitude and landform. In addition, the 3D landscape may be understood more intuitively than altitude and landform encodings in traditional maps (Rase 2003, Meng 2003). To illustrate the importance of a combined visualisation of data and landscape a very early (2D) example is shown in figure 1.5. The cartographer Jacob Scheuchzer represented the prevailing religion of small cities and towns of the Toggenburg area in eastern Switzerland in 1710, towards the end of about 300 years of religious war between Catholics and Protestants. He could have done so on a 2D map showing the towns and main roads. However, he decided to depict also the hilly landscape of the area. At that time the mountains were still a dramatic and very symbolic element of nature (Ambroziak & Ambroziak 1999). Additionally, the spread and adoption of a new religion is also largely dependent on the location or seclusion of a town, the valleys, hills and passes and thus on the three dimensional landscape they lay in. Thus, it may be easier to understand why a secluded village was still catholic while the surrounding (but separated by hills) villages took on the protestant religion when the three dimensional landscape is visualised together with the religion of each village.

Figure 1.5.: Map of the Toggenburg area showing the religion of the villages and localities (Ambroziak & Ambroziak 1999)

In section 1.2.1 it is outlined that combinations of data displays with virtual environments are easily created with the help of available tools and virtual globe technologies and that they are also increasingly popular. Wood et al. (2005) point out that most 3D geovisualization approaches are technology driven rather than theory driven. They state that there "is a risk that, for example, a virtual walk-through or fly-by will meet society's zeitgeist more than cognitive cartographic requirements." (Wood et al. 2005, p. 306) and remark that we still know too little about when and how 3D visualisations can be used appropriately and effectively. The need for more research in 3D geovisualization and the development of 3D related theory is also reflected in the various research agendas examined in section 1.2.3.

Studies evaluating 3D displays are discussed in section 1.2.2. The main areas of research are application areas and tasks where 3D is appropriate, the differences between 2D displays, 3D desktop-based virtual environments and immersive environments such as the CAVE and the interaction with virtual environments. While some studies report better performance of 3D visualisations, others do not find a difference between 2D and 3D or report that 2D visualisations perform better. Overall, the results of many studies seem dependent on the exact study design including tasks, information displayed and display characteristics. What is generally missing is the implementation and evaluation of data displays within virtual environment and research into how appropriate such displays are for visual data exploration and analysis. For appropriate visual encodings of data within virtual environments it is sensible to start from established cartographic rules and guidelines for 2D displays, such as thematic maps, developed over many years (e.g. Slocum et al. 2005). Kraak (1988) stated that the general cartographic theory should be applicable to 3D visualisations if they are kept relatively simple but asks for further comparison between 2D and 3D maps to prove this. Not all cartographic 2D theory may be valid for 3D likewise (Slocum et al. 2001). For example, in perspective 3D displays every displayed object does not only vary in size with the value it actually represents but also with the depth cue it provides (e.g. objects in the background of the scene are smaller than objects in the foreground). Thus, the display of data values within virtual environments needs testing and evaluation to find appropriate visual encodings supporting effective perception of values and patterns as needed for data exploration and analysis. The results of such evaluations may help to establish new guidelines and rules especially for 3D geovisualization. Additionally, displays of abstract data within virtual environments need to be evaluated to understand and to judge the value of such displays in different application areas and for different tasks (Slocum et al. 2001). To gain this holistic view a single research method is too limited. A combination of different research methodologies and stages which are designed driven by specific visualisation and application characteristics is needed. Such a methodological framework is supposed to be able to evaluate the visualisation of quantitative data within virtual environments as holistically as asked for and, additionally, may be applicable to future evaluations of visualisation technologies if it proves appropriate.

1.4. Aims and research questions

On the basis of the above introduction, review of research findings and research agendas as well as the problem statement the research reported here aims to:

- Increase understanding of the use of desktop-based virtual environments with a focus on the graphical representation of quantitative data through abstract symbols or graphics.
- Relate perceptual/cognitive studies and case studies in specific application areas by evaluating the use of desktop-based virtual environments for analysing quantitative data in a range of experimental and applied settings.

1.4.0.1. Research questions

- Are methods of quantitative data representation derived from literature about 2D representation methods appropriate in desktop-based virtual environments in an experimental context?
- Are these methods appropriate for more complex visualisation tasks involving landform?
- Are these methods appropriate in applied scenarios with real tasks and data experts for the visualisation of quantitative data in relation to the landform?

1.4.0.2. Hypotheses

The following main hypotheses are tested:

- Typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative values in virtual environments when using appropriate representation methods.
- Typical users are able to relate multivariate data representation displayed in virtual environments.
- Typical users are able to relate appropriate representations of quantitative data to landform.
- Data and task experts gain insights into their data sets when appropriately displayed in virtual environments depicting the landscape to which the data relates.

Table 1.1 assembles all research questions, hypotheses and propositions tested and evaluated at the different stages of the research.

For a quick overview of the results of the research in regard to the proposed aims, objectives and hypotheses jump to table 5.1 in chapter 5 which shows the same three top table rows of aims, research objectives and main hypotheses with indication of acceptance (✔) or rejection (✘) of each. Additionally, the tables 4.7 (stage I), 4.8 (stage IIa), 4.9 (stage IIb) and 4.10 (stage III) in chapter 4 show the research hypotheses, propositions and question of each stage also with indication of acceptance or rejection.

aims	To increase understanding of the use of desktop-based 3D virtual environments with a focus on the graphical representation of quantitative data through abstract symbols or graphics. To relate perceptual/cognitive studies and case studies in specific application areas by evaluating the use of desktop-based 3D virtual environments for analysing quantitative data in a range of experimental and applied settings.			
research objectives	Are methods of quantitative data representation derived from literature about 2D representation methods appropriate in desktop-based virtual environments in an experimental context?	Are these methods appropriate for more complex visualisation tasks involving landform?		Are these methods appropriate in applied scenarios with real tasks and data experts for the visualisation of quantitative data in relation to landform?
	stage I	**stage IIa**	**stage IIb**	**stage III**
main hypotheses	Typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative values in virtual environments when using appropriate representation methods.	Typical users are able to relate multivariate data representations displayed in virtual environments. Typical users are able to relate appropriate representations of quantitative data to landform.		Data and task experts gain insights into their data sets when appropriately displayed in virtual environments depicting the landscape to which the data relates.
hypotheses H / propositions P / questions Q for each stage	H I.I: Users are able to identify the taller of two bars in a 3D desktop virtual environment as well as they can in static 2D graphics. H I.II: The effectiveness of estimating differences between two bars is not significantly different in the 2D and 3D settings. H I.III: The efficiency of task completion is improved by the use of a reference grid (frame) (in the 2D and 3D settings). H I.IV: The efficiency of estimating absolute values from bar lengths (with a reference frame) is not significantly different in the 2D and 3D settings.	H IIa.Ia: Tasks where the reference set is location are solved more efficiently in the 2D representation. H IIa.Ib: Tasks where the reference set is location are solved more effectively in the 2D representation. H IIa.IIa: Tasks where the reference set is altitude are solved more efficiently in the 3D virtual environment. H IIa.IIb: Tasks where the reference set is altitude are solved more effectively in the 3D virtual environment. H IIa.IIIa: Tasks that include combined reference sets (location and altitude) are solved more efficiently in the 3D virtual environment. H IIa.IIIb: Tasks that include combined reference sets (location and altitude) are solved more effectively in the 3D virtual environment.	H IIb.Ia: Insights into the relationship between data and location/landcover/etc. are more often and more efficiently reported in the 2D representation. H IIb.Ib: Insights into the relationship between data and altitude/landform are more often and more efficiently reported in the 3D virtual environment. H IIb.II: The complexity and plausibility of insights reported do not vary significantly between the 2D representation and the 3D virtual environment.	P III.I: Displays of quantitative data within virtual environments are more appropriate for the visual analysis of the data if the data and landscape (2D/3D) relationship is more important in an application. P III.II: A visual combination of data and landscape allows more efficient and/or more effective finding of insight into the data. Q III.III: Can differences in appropriateness of the visualisation for data analysis be related back to different characteristics of the evaluated application settings?

Table 1.1.: Overview of explored and tested aims, research objectives and hypotheses

1.5. Summary of contributions

The main contributions of this research are summarised below.

- An appropriate visual encoding for quantitative data in desktop-based 3D virtual environments was derived based on 2D cartographic theory and knowledge about depth cues and the human visual system (2D bars or bar charts with reference frames on billboards). Typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative data in virtual environments when represented as 2D bars on billboards. Participants preferred single bars or simpler bar charts displaying a maximum of 3-4 variables. The reference frames which are beneficial for comparing pairs of non-overlapping values added to the visual complexity in more data dense scenes. The degree of overlap and the number of variables that can be interpreted will depend upon the setting, the data set as well as the needs and capabilities of its users. (⇒ sections 2.3.1 and 5.2.1.1)
- Initiation of the development and testing of an XML based process for the semi-automatic creation of quantitative data displays within virtual environments. The process is available under a Creative Commons license. (⇒ section 2.3.2)
- Design, implementation and evaluation of a methodological framework bridging different quantitative and qualitative research stages from 'in vitro' to 'in vivo' for the evaluation of a visualisation technique. Methodologically bridging between experimental 'in vitro' and case study based 'in vivo' research methods is appropriate as the results of each stage can inform the design of the following stages. Additionally, the outcomes of later stages lead to re-evaluation of or different interpretation of earlier results. In combination a holistic view of the appropriateness of the evaluated display method is gained. (⇒ sections 2.1.0.3 and 5.2.2)
- Testing of the appropriateness of the devised visual encoding for quantitative data in virtual environments in a series of studies with increasingly complex data sets and tasks from controlled experiments to case studies in application areas with informed participants and data experts. The users are able to relate multivariate data represented in virtual environments in general and also to altitude and landform but the 3D environment does not especially support this. Variation in the results between the data analysis in either the 2D or the 3D visualisation is insignificant. (⇒ sections 2.5 and 5.2.1.2)
- The results demonstrate that the varying characteristics of data sets strongly influence the appropriateness of either the 2D or the 3D visualisation for data analysis. Tentative recommendations are given about when to sensibly use either type of visualisations. A combination of both visualisation types should be researched further. (⇒ sections 5.2.1.3 and 5.5)
- It was found that neither the informed participants nor the data experts are used to data analysis in relation to altitude and landform. However, they are able to do it when guided by tasks but report difficulties in task completion. (⇒ sections 5.2.1.4 and 5.2.2)
- Two sets of recommendations for the use of quantitative data displays in virtual environments and the employed evaluation methods based on the findings and experiences of this research are provided. (⇒ sections 5.3.1 and 5.3.2)
- The results have implications on future evaluations of visualisation techniques. The strong influence of the varying characteristics of data sets used in this evaluation and the variation even

amongst informed participants requires quick 'validation' methods. Additionally, the findings of this research should guide future design and evaluation processes of visualisations techniques and prototypes. (⇒ section 5.4)

- The results of this research close a gap as displays of quantitative data within virtual environments were proposed and/or developed mainly without evaluating appropriateness so far. (⇒ sections 1.2.1, 1.2.2 and 1.2.3)

Preliminary results and findings of this research were presented and discussed at different conferences and workshops:

- GIScience 2006, Münster (Bleisch, Dykes & Nebiker 2006)
- MPhil-PhD upgrade Seminar 2007, London (Bleisch 2007)
- ICA Commission on Visualization and Virtual Environments Workshop 2007, Helsinki (Bleisch, Dykes & Nebiker 2007)
- XXI ISPRS Congress 2008, Beijing (Bleisch & Nebiker 2008)
- Doktorandenkolloquium FHNW & DLR 2008, Muttenz (Bleisch 2008)
- GISRUK 2009, Durham (Bleisch, Dykes & Nebiker 2009*b*)
- Workshop Human Aspects of Visualization, INTERACT 2009, Uppsala (Bleisch, Dykes & Nebiker 2009*a*)
- 3. Anwendertreffen GIS in Nationalen Naturlandschaften, 2. Workshop GIS within the network of alpine protected areas, 2009, Zernez (Bleisch 2009)
- 24th International Cartography Conference 2009, Santiago de Chile (Bleisch, Burkhard & Nebiker 2009)
- geosuisse Winterveranstaltung 2010, Zürich (Bleisch 2010*a*)
- Workshop on Methods and Techniques of Use, User and Usability Research 2010, London (Bleisch, Dykes & Nebiker 2010*a*)
- GIScience 2010, Zürich (Bleisch, Dykes & Nebiker 2010*b*)
- Kartographisches Kolloquium 2010, Karlsruhe (Bleisch 2010*b*)
- FHNW HABG BrownBag Lunch 2010, Muttenz (Bleisch 2010*c*)

The results and findings are reported in this thesis write-up and published in The Cartographic Journal (Bleisch et al. 2008, ➪ Appendix A).

2. Methods

< Chapter 1. Introduction

The aims and research objectives of this study ask for an evaluation of the appropriateness of visual combinations of quantitative data representations with virtual environments (2.1). First a suitable visual encoding of quantitative data values is derived from literature about 2D cartography and knowledge about the human perception and depth cues in desktop-based 3D displays. Based on this, it is decided that displaying the data values as 2D bars on billboards with reference frames in virtual environments is most appropriate (2.3.1). The data displays in virtual environments are experimentally compared to more traditional 2D representations with data displays. The first experiment showed that the reference frames are helpful for the 3D displays but not the 2D displays. Thus, after the first study, the reference frames are no longer included in the 2D representations. The different 2D and 3D displays are created using a semi-automatic XML based process (2.3.2).
The holistic evaluation asked for in the research goals cannot be achieved by employing a single research method (2.1.0.3). While on one hand hypotheses need to be confirmed, which typically asks for quantitative experiments, on the other hand the use of desktop-based 3D virtual environments for the display of quantitative data shall be explored and a deeper understanding gained which is typically done employing case study methods. Thus, a methodological framework is designed which implements a mixed methods approach to complement the strengths and overcome the weaknesses of any single research method (2.1.0.4). A 'bridge' of four research stages from 'in vitro' experimental studies to 'in vivo' case studies is identified to accommodate increasingly complex data sets and tasks (table 2.1). The experimental stages are evaluated with the help of informed users (2.4). Data experts participated in the case studies of stage III (2.5.4). For the research stages, different tasks are defined based on the functional view of data and tasks which allows separating between tasks referring to different reference sets (2.2). To measure appropriateness (2.1.0.2) the task answers are evaluated either quantitatively in regard to task accuracy and task performance time (2.6.1) or qualitatively using insight evaluation methods in combination with quantitative task performance time analysis. Insights are evaluated in regard to complexity, plausibility, reference type, word usage and user confidence (2.6.2). The case study data of stage III is evaluated by compiling and analysing cross-case 'thick matrix displays' and following a replication strategy for pattern detection (2.6.2.5). Each data set collected at the four research stages is handled according to its characteristics to prepare it for data analysis (2.6.3).
Stage I implements a controlled experiment to evaluate if typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute data value changes on the 2D bars on billboards within virtual environments (2.5.1). 26 informed participants performed two elementary tasks of identifying the taller of two bars and comparing bar lengths in a selection of 2D and 3D representations of combinations of two bars. Task performance time was recorded. The experiment was conducted using a 2x2 factorial within-subject design randomising the order of the bar combination assignments. The four factors are visualisation type (static 2D or interactive 3D virtual environment) and data graphic type (bars with and without reference frames).

Stage IIa employs an experimental setting to evaluate if typical users are able to relate data representations displayed in virtual environments and if they are able to relate the data representations to altitude and landform (2.5.2). Eight single variate data sets representing deer tracking data during different hours of the day during summer and winter months are used to prepare 16 visualisations, each data set in 2D and 3D. The four data sets representing the winter months were spatially transformed to a nearby area to allow for evaluation of the influence of different settings. 34 informed participants answered seven tasks in each of two 2D representations and two 3D representations which were assigned to them in a balanced order. The seven tasks are of varying complexity (elementary and synoptic) and have different reference sets. Two tasks refer to location, three tasks refer to altitude and two tasks refer to location and altitude in combination.
Stage IIb implements an experimental setting to evaluate if typical users are able to relate multivariate data displays in virtual environments to altitude and landform (2.5.3). Two data sets consisting of the aggregated four summer and four winter data sets of stage IIa are used to prepare four visualisations, each data set in 2D and 3D. The data set representing the winter months was spatially transformed to a nearby area to allow for evaluation of the influence of different settings. An insight reporting method based on one synoptic task was employed to avoid the influence the seven tasks in stage IIa had on the task answers. The task refers to location, altitude and the attribute time of the day. 38 informed participants reported insights and judged their confidence in each insight during analysing one 2D and one 3D representation, each for half an hour. The visualisations were assigned to them in a balanced order.
Stage III employs a descriptive and explorative case study approach to evaluate if the display of quantitative data within virtual environments is appropriate in applied scenarios with real tasks and data experts (2.5.4). To enhance representativeness by allowing for cross-case analysis a multiple case sampling strategy selecting diverse cases which vary in the dimensions of interest (e.g. data density, data covered area and topography or relation of data and landscape) is followed. Three case studies 'Brienz', 'Literature Atlas' and 'Deer SNP' are selected based on their data characteristics and the availability of the data experts. To compare the data experts to the informed participants who took part in the other three stages, the data experts worked through the tasks of stage I. According to the case study protocol, each data expert provided a data set and context information for the case before data visualisations within virtual environments were prepared. The data experts were then visited in their workplace to discuss their use of the visualisations. The data collection process was completed with a debriefing one week after the visit.

Chapter 3. Data analysis and results >

2.1. Methodological overview of the research

The aims and research objectives of this study (section 1.4) ask for an evaluation of the appropriateness of visual combinations of quantitative data representations with desktop 3D virtual environments. To be able to do this, the research is dependent on appropriateness in regard to several aspects and characteristics of combined visualisations and also evaluation methods. The study needs suitable ...

1) ... representations of the quantitative data.

2) ... measures of appropriateness, including task definitions and a benchmark for comparison.

3) ... evaluation methods.

Each of the following sections introduces one of these aspects and references later sections of this chapter where more details can be found. In combination, these first sections give a methodical overview of the whole study.

2.1.0.1. Representations of quantitative data

The quantitative data needs to be suitably visualised within desktop-based 3D virtual environments. Section 2.3.1 details how a suitable visual encoding for the data is derived from the literature about 2D cartography, knowledge about the human perception and the depth cues existent in desktop-based 3D virtual environments. Based on the 2D cartographic knowledge the data is then displayed as 2D bars or bar charts (combinations of single bars) on billboards within the virtual environments.

The representation of the virtual environments (section 1.2.0.3) is based on the de-facto standard which is used by most available virtual globe technologies - a digital elevation model draped either by a map or high resolution ortho imagery (sections 2.5.1-2.5.4 for details about the data sets) displayed on the 2D computer screen. If not specifically mentioned the virtual environments provided by Google Earth (Google 2010) are used as base for the evaluations in this study. This technology also provides a de-facto standard for interaction (3D navigation). The informed users of this study (section 2.4 for details about the participants) are assumed to be at least as familiar with virtual globes and the interaction with them as two-thirds of the general population (Bartoschek & Schönig 2008). In addition, each evaluation starts with time for introduction and getting used to the technology and the tasks (sections 2.5.1-2.5.4 for details about the research process).

2.1.0.2. Measures of appropriateness

The appropriateness of visualisations is often measured based on efficiency and effectiveness (e.g. Mackinlay 1986, van Wijk 2005). Cowie (1989, p. 386) defines efficient as ”producing a satisfactory result without wasting time or energy”. Effective is defined as ”producing the intended result” or being ”fit for service or work” (Cowie 1989, p. 386). Jobst & Germanchis (2007, p. 222) note that ”effectiveness regards aesthetic concerns as well as effective information acquisition [and] [. . .] interface use [. . .]” (originally stated by Mackinlay 1986). van Wijk (2005, p. 79) colloquially says that a ”visualization should do what it is supposed to do, and has to do this using a minimal amount of resources.” But how to measure it? Johnson (2004) even lists the quantification of effectiveness as one of the top scientific visualisation research problems. Typically, task completion time and success or error rates are measured (e.g. Chen & Yu 2000, Tobon 2005, Tory et al. 2006). Zhu (2007) proposes a more detailed view of visualisation effectiveness based on the three principles of accuracy, efficiency and utility. Accuracy defines the relationship between data and visualisation which can be measured through the number of interpretation errors. Efficiency defines the relationship between visualisation and users and is, for example, measured through records of task completion times. The third principle of utility defines the relationship between visualisation and tasks and could according to Zhu (2007), for example, be measured by how well a number of specific task goals are achieved.

In the context of this research efficiency and effectiveness of the visualisations are evaluated. Efficiency is measured by recording and analysing task completion times. The measures of effectiveness are based on the concepts of accuracy and utility. Accuracy of the quantitative data representations is measured through interpretation error rates (section 2.5.1). However, this is only possible in experimental settings with tasks having a true or exact answer which can be evaluated. The research questions of this study also go further and an evaluation of utility by measuring how well task goals are achieved is needed. North (2006) proposes to evaluate and measure insights generated by the visualisation as ”the purpose of visualization is insight” (North 2006, p. 6). He defines insights as being complex, deep, qualitative, unexpected and relevant. Insights thus have no exact true or false answer and are more difficult to evaluate than error rates. In this research insights are evaluated by rating the complexity, plausibility and recording the user confidence (Rester, Pohl, Wiltner, Hinum, Miksch, Popow & Ohmann 2007) of each insight (section 2.6.2.1).

Task de nitions: Measuring the appropriateness of a visualisation, based on efficiency (task completion time) and effectiveness (task accuracy and utility based on task goal achievement) as discussed above, requires suitable task definitions. Thus, tasks need to be manageable in a certain time, need a precise result for evaluating the accuracy or guide the participants in finding insights. A range of typical geovisualization task definitions have been reviewed and the functional view of data and tasks defined by Andrienko & Andrienko (2006) proved most suitable as it allows the separation of tasks with different reference sets such as location or altitude (section 2.2 for details about task definitions) as asked for in the research hypotheses (section 1.4.0.2).

Benchmark for comparison: Besides defining suitable tasks, comparing the performance of quantitative data displays in 3D virtual environments to the performance of a potentially more familiar 2D visualisation of the same data and background information may be helpful as 2D visualisations for the spatial analysis of data are more widespread. Thus, the evaluations include comparisons between data displays in 3D virtual environments and displays of the same data on 2D representations (map or ortho imagery) of the same environment (section 2.3.2 for details about the 2D and 3D representations).

2.1.0.3. Evaluation methods

Various evaluation methods with their characteristic strengths are available for evaluations in geovisualization and cartography (e.g. Slocum et al. 2001, Montello 2002, Marsh 2007). Traditionally, empirical quantitative techniques, such as controlled experiments, are used to establish knowledge about cognitive responses to maps (Montello 2002). Experiments usually involve large numbers of users, with little contextual information and thus no (or few controllable) influencing factors. In applied research settings very few users are typically involved and qualitative approaches are employed. Much contextual and tacit knowledge influences and enriches these studies and many influencing factors exist that we cannot or do not want to control (Yin 2003). Research in geovisualization is in many cases done by employing controlled experiments (e.g. Bair & House 2007, Fabrikant, Montello & Mark 2006). Case studies or applied settings are used when assessing implementation or usability issues (e.g. Brooks & Whalley 2008, Koua, MacEachren & Kraak 2006). Other approaches use experimental settings for evaluating the appropriateness of different visualisation types to ease understanding of a data set and to gain insight into it (Rester et al. 2007).

Considering the posed aims and research questions (section 1.4) and their complexity it is evident that they cannot be answered using a single research method. While on one hand hypotheses need to be confirmed, which typically asks for quantitative experiments, on the other hand the use of desktop-based 3D virtual environments for the display of quantitative data should be explored and a deeper understanding gained which is typically done employing case study methods. To enable a research project with such differing types of question, a mixed methods research concept combining different methodological approaches is applied. The mixture of methods shall result in ”complementary strengths and nonoverlapping weaknesses” (Johnson & Onwuegbuzie 2004, p. 18). Figure 2.1 shows the most important concepts and factors that influence the research reported here. Employing a combination of research methods makes it possible to ’bridge’ between ’in vitro’ and ’in vivo' research methods. The following section gives an overview of the employed research methods for each research stage.

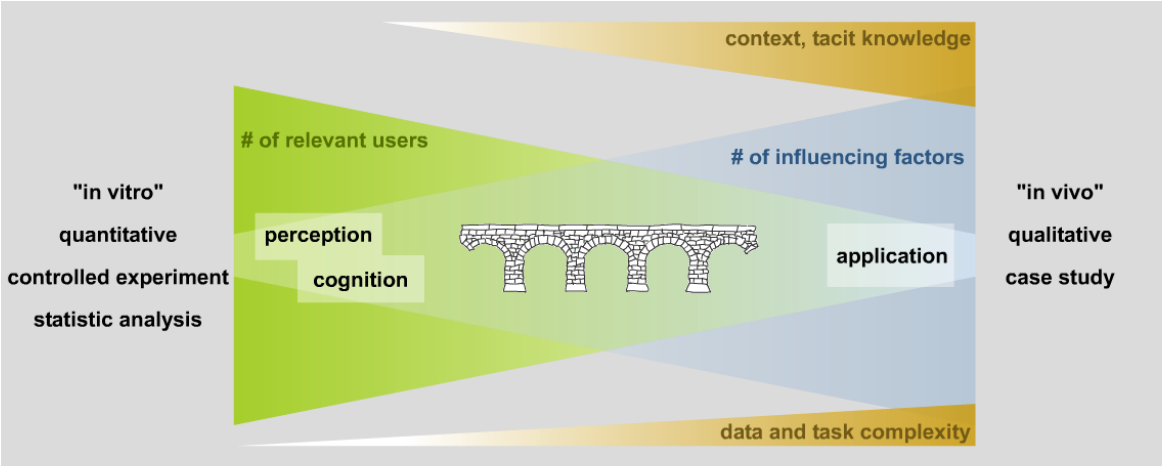


Figure 2.1.: The bridge between the two sides of 'in vitro' and 'in vivo' research approaches

2.1.0.4. Overview of research stages

To test the display of quantitative data in virtual environments a series of research stages along the bridge from 'in vitro' to 'in vivo' research (figure 2.1) are identified. They correspond to the research objectives defined in section 1.4 and are determined as important research steps while taking into account the diverse influences such as increasing context as well as increasing data and task complexity from an experimental setting to real world settings. In stage I, hypotheses about the appropriateness of data graphics within virtual environments is tested 'in vitro' employing controlled experiments. This is done by employing a somewhat psychophysical approach measuring responses to a known stimulus (bar length). Such approaches were very popular in cartography and map design research some years ago (Montello 2002). In relation to the series of research methods conducted along the bridge it makes sense to first test if the selected display method is suitable for basic tasks and data before evaluating how it behaves with more complex tasks, data and in real world contexts. Montello (2002, p. 295) reminds us that psychophysical studies are characterised by their focus on the perception of isolated symbols without integrating any context or cognitive tasks but that "such low-level tasks are an essential precondition for seeing anything on a map."

On the other end of the bridge, the appropriateness of displays of quantitative data in virtual environments is explored in real world settings 'in vivo'. Typically in real world settings, a small number of experts work with large data sets. Their use of complex visual tasks to better understand the data may best be researched using a case study method (Yin 2003) which takes into account and values the many influencing factors of such a complex setting and explores issues such as characteristics of potential applications and users. Thus, stage III employs different case studies of real world settings. This, compared to stage I, includes high level cognitive tasks, complex data sets, data experts and all the context information of a real world setting. Hutchins (1995) even proposes that cognition can only validly be assessed and evaluated 'in the wild' or in real world settings.

Stage I and stage III of this research span the breadth of research in geovisualization. Slocum et al. (2001, p. 68), in their research agenda, identify that "we need to examine the effectiveness of geovisualization methods, both in the traditional laboratory setting (where they are developed) and in the 'real world' (where they are actually used)." However, there is a gap or at least a big step between evaluating the display of quantitative data in virtual environments only at the two ends of the continuum. To really bridge, and not only to jump, research stage II is identified which is positioned between the 'in vitro' controlled experiment of stage I and the 'in vivo' case studies of stage III. This stage II is subdivided into the stages IIa and IIb as the steps in increasing data and task complexity would be to large to be comparable from one stage to the next stage along the bridge without subdivision. In two experimental settings first the task complexity (in stage IIa) and then also the data complexity (in stage IIb) is increased to build a bridge from the stage I laboratory setting to the highly complex real world settings of stage III. Thus, it is tried to keep the steps between the research stages as minimal as possible to ensure better comparability and integration of the results along the bridge. In table 2.1 the characteristics of each research stage are compiled for an overview. Details about each research stage are available in sections 2.5.1 (stage I), 2.5.2 (stage IIa), 2.5.3 (stage IIb) and 2.5.4 (stage III).

To accommodate the diverse research stages the study employs a sequential mixed methods approach (Johnson & Onwuegbuzie 2004). Each of the identified research stages uses its own research method(s) and the results are used to inform the following stage(s). While the collected data of each stage is mainly analysed separately, it is evaluated integratively across all stages for certain aspects. Additionally, the results and findings of all stages are integrated in the discussion. This allows conclusions to be drawn forward and backward along the research stages. While in the forward process findings from an earlier stage inform later research stages backward inferences are done by revisiting and re-evaluating earlier collected data and findings in the light of later findings. The main purposes for conducting mixed methods research in this study is complementarity (to elaborate and enrich findings from one method with results from another), expansion (aiming for breadth) and triangulation (seeking convergence of findings) (Greene, Caracelli & Graham 1989) in accordance with the research aim of increasing understanding of the use of desktop-based virtual environments with a focus on the graphical representation of quantitative data through abstract data graphics.

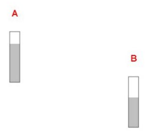
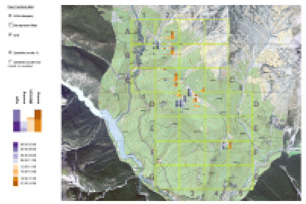
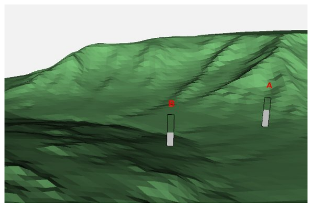
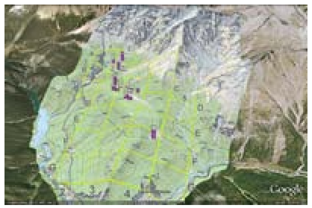
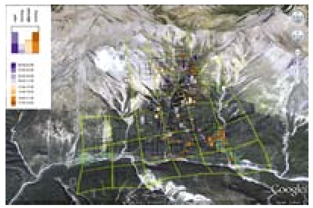
	stage I (section 2.5.1)	stage IIa (section 2.5.2)	stage IIb (section 2.5.3)	stage III (section 2.5.4)
data	univariate (20 data sets, random values)	univariate (8 data sets, # of deer visits per location)	multivariate (2 data sets, # of deer visits per time of the day and location)	multivariate (1-2 data sets per case)
data graphics (section 2.3)	single bars	single bars	bar charts	bar charts
task level (section 2.2)	elementary	elementary & synoptic	synoptic	(elementary &) synoptic
tasks (section 2.2)	two tasks (lookup, comparison)	seven tasks referencing either location, altitude or both	one task referencing location and altitude	reported by data experts
main methods (section 2.5)	experiment, digital questionnaire (face-to-face)	experiment, digital questionnaire (online)	experiment, digital form for insight reporting (face-to-face)	case studies, semi-structured interview, digital questionnaire (face-to-face and email)
participants (section 2.4)	2x2 factorial within-subject design 26 final semester geomatics students	balanced within-subject design 34 final semester geomatics students, former students MGI City University and staff FHNW, City University and UZH	balanced within-subject design 38 final semester geomatics students	multiple case design (diverse cases) 3 case and data experts
collected data	time, identification of taller bar, bar length differences, comments	time, task answers, confidence ratings, comments, think-aloud protocols	time, reported insights, confidence ratings, comments	data and case description, interview protocol, comments
data analysis (section 2.6)	quantitative (comments qualitative)	quantitative and qualitative	quantitative and qualitative	qualitative
example 2D representation (screenshots)				(no 2D visualisations in stage III)
example 3D virtual environment (screenshots)				
	(for all screenshots see figures 2.8 and 2.9)	(for all screenshots see figure 2.10)	(for all screenshots see figure 2.11)	(for all screenshots see figure 2.12)

Table 2.1.: Overview of the characteristics of each stage and the differences between the research stages (including references to the sections and figures where details are available)

2.2. De ning tasks

The definition of tasks is a complex issue in geovisualization evaluations (Tobon 2005). Searching the literature yields numerous task definitions and also numerous studies which apply them. Interaction taxonomies (e.g. Yi, Kang, Stasko & Jacko 2007) often combine navigation, data display manipulation and tasks. In the context of this research it is attempted to separate them. The navigational part of interaction is kept as standard as possible by using the functionality offered by established 2D SVG displays and virtual environments mainly based on Google Earth (section 2.1.0.1) with which informed participants (section 2.4) are supposed to be minimally familiar. Data manipulation is not a goal of the research and thus not possible. The main task which shall be evaluated in this research is the exploration of data sets in relation to the landscape. Thus, suitable task definitions are required which can appropriately reflect the goals aimed for. The difficulty with most task definitions is that they are either rather tool based, for example, overview, filter or details-on-demand (Shneiderman 1996), computational as, for instance, retrieve value, find extremum or sort (Amar, Eagan & Stasko 2005), or mainly concentrated on the analysis of the data and do not seem to be able to include the relationship of data and landscape into the analysis. Examples for data analysis tasks include identify, compare or categorise as defined in task taxonomies by Wehrend & Lewis (1990), Keller & Keller (1993) or Zhou & Feiner (1998). These task definitions are often used for evaluating geovisualizations (e.g. Koua et al. 2006, Nekrasovski, Bodnar, McGrenere, Guimbretiere & Munzner 2006, Morse, Lewis & Olsen 2000). Additionally, adaptations and/or extensions of these task taxonomies are proposed and implemented by a number of researchers (e.g. Valiati, Pimenta & Freitas 2006, Xiang, Chau, Atabakhsh & Chen 2005, Ogao & Kraak 2002).

The functional view of data and tasks defined by Andrienko & Andrienko (2006) is based on the distinction between characteristic and referential component of data. The characteristics of data are, for example, the measurement values or observations and the referential component specifies the context, such as a place. Thus, a data set can be viewed as a set of links between references (R) and characteristics (C) as symbolically represented in figure 2.2.

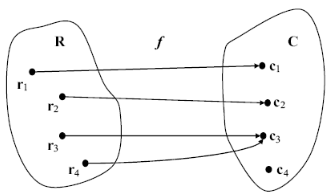


Figure 2.2.: Functional view of a data set. A set of references (R) linked to a set of characteristics (C) (Andrienko & Andrienko 2006, p. 7)

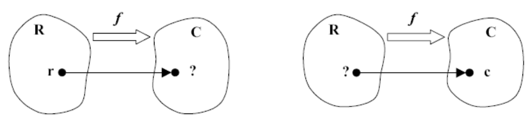


Figure 2.3.: Two elementary tasks represented schematically on the basis of the functional view of data in figure 2.2 (Andrienko & Andrienko 2006, p. 8)

Based on this data definition the data analysis tasks can be defined in terms of the two data components characteristics and references. Andrienko & Andrienko (2006) differentiate between elementary tasks dealing with single elements of data (e.g. single characteristic or reference) or synoptic tasks dealing with a set of references or characteristics. The task is then defined as having a target (what needs to be known) and having constraints (what is already known and has to be taken into account). Two elementary tasks are schematically represented in figure 2.3. On the left hand side the reference is known and we want to know the characteristics of the data at that reference, for example, the measured value at a specific location (direct lookup). And on the right hand side the reversed case, we know a specific characteristic, for example an observed value, and want to know at what reference (where) it was observed (inverse lookup). Based on the same logic Andrienko

& Andrienko (2006) define lookup, comparison and relation-seeking tasks on the elementary and synoptic level.

For comparisons of results between different visualisation evaluation studies, it is helpful that most of the data analysis tasks defined by other researchers (e.g. Wehrend & Lewis 1990, Keller & Keller 1993, Zhou & Feiner 1998) as detailed above, can also be defined using Andrienko & Andrienko's (2006) functional data and task definition.

2.3. Data representation

2.3.1. Data graphics design

Data visualisation displays are well researched in 2D and recommendations exist about what abstract symbolism works best for what type of data. For example, it is common to use the length and area of abstract symbols to effectively show quantity (Tufte 2001, Cleveland & McGill 1984). Thus, typical displays for the comparison of different data values include bars (e.g. Few 2009), parallel coordinate plots (e.g. Inselberg 2008) or graduated point symbols such as circles of varying size (e.g. Flannery 1971).

Data displays in desktop-based virtual environments are subject to depth cues. The perception of 3D virtual environments projected onto 2D screens needs monocular depth cues such as linear perspective, size gradient, occlusion or structure from motion (Ware 2004) which trick us into seeing the depicted environment in 3D. In the real world, most people are aided in their judgment and perception of the environment through their ability to see in stereo. However, this ability is limited to the immediate vicinity. For objects that are more distant and sometimes also for closer objects, we rely on depth cues. Ware (2004, p. 289) even argues that "stereoscopic depth can play no role at all at distances beyond 30m". Thus, we may be well-trained in judging objects according to depth cues rather than seeing them stereoscopically. Ware (2004, p. 262) paraphrases Hagen (1974) stating "when we perceive pictures of objects, we enter a kind of dual perception mode. To some extent, we have a choice between accurately perceiving the size of the depicted object as though it were in a 3D space and accurately perceiving the size of the object at the picture plane." Rock (1998) explains that in the real world the user is able to judge the size of objects (relative differences) because the position in the perceived 3D environment is known (size constancy).

All elements displayed in a desktop-based 3D virtual environment are distorted to some degree as they provide depth cues such as size gradient. Or in other words, for example, graduated data symbols in a 3D virtual environment do not only change according to the values they represent but also depending on their position in the landscape (objects further away are displayed smaller than objects nearby). Nevertheless, most users, even scientists, interact with three-dimensional displays on their 2D desktop computer screen as more immersive environments such as the CAVE are not readily available and expensive (Slocum et al. 2001). Thus, we need to find representations for quantitative data which despite the influence of the depth cues communicate the data values appropriately. Already Gahegan (1999, p. 289) identified "the management of perceptual anomalies due to visual combination effects" as one of four barriers to the development of effective exploratory visualisation tools.

Looking at the two examples of data displays in virtual environments in figures 2.4 and 2.5 we discover a number of problems which arise from the depth cues and the possibility to freely change the viewpoint in an interactive virtual environment. In figure 2.4 the data values (children under five mortality rate per country) are displayed as graduated circles. Is your judgement of size difference, and thus value difference, between the two countries (indicated by the red and green arrows) the same from both viewpoints? Circular symbols in virtual environments do not provide any clue as to how big the influence of the depth cue size gradient is. But judgements may be helped by the surrounding information, such as size gradient of the labelling. Figure 2.5 shows varying data values with the length of hexagonal 3D volumes. As the volumes are positioned statically the viewpoint can be changed to look at them from a top view. From there the length of the volumes is no longer readily visible and the creator of the display needs to additionally use colour to convey the differences in values. According to Tufte (2001) it is bad practice to use two visual variables to communicate one data value. Similar examples which use some sort of 3D symbolisation and additionally colour to communicate the differing values are found in figures 1.3a, 1.3c, 1.3g, 1.3h, 1.3i and 1.3k. Double encoding of data values in virtual environments is common (bad) practice. Tufte (2001) pointedly terms it a waste of 'ink'.

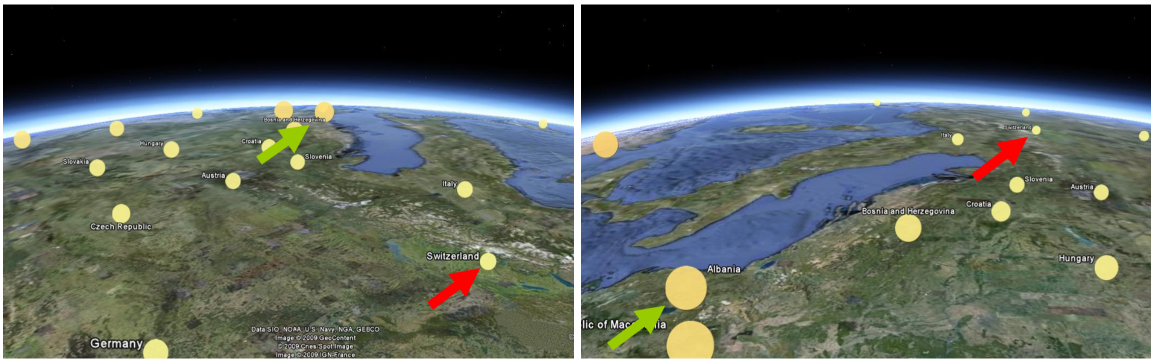


Figure 2.4.: Compare Switzerland (red arrow) with Albania (green arrow) from two different viewpoints (children under five mortality rate, circular symbols in Google Earth; display created with thematicmapping.org/engine (Sandvik 2010))

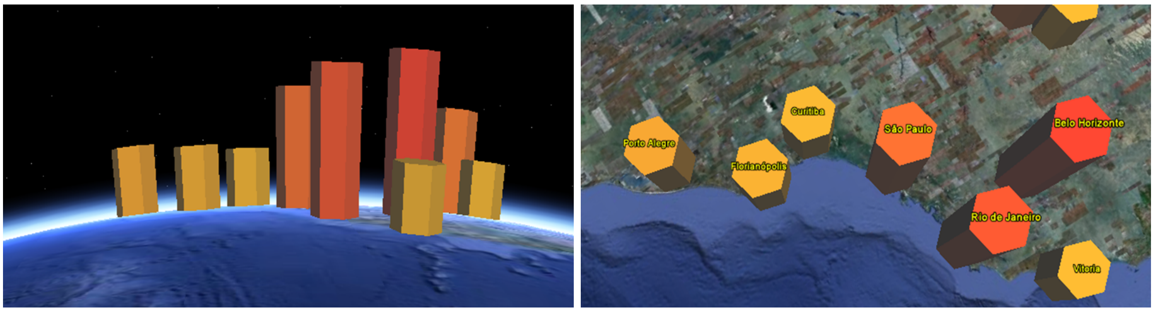


Figure 2.5.: Hexagonal 3D bars in Google Earth - side and top view (display created with GE-graph (Sgrillo 2010))

Both of the above mentioned difficulties, looking at 3D symbols from an unfortunate angle and circular symbols not providing a reference for size judgement, can be overcome by using displays of 2D bars on billboards. Billboards are 3D computer graphic elements consisting of a raster image which is rendered onto a transparent 2D polygon (mostly rectangular) (Akenine-Möller & Haines 2002). The billboard is positioned fix within the virtual environment, at the location where the displayed data values relate to, but it rotates along two axes to have its graphic face always aligned vertically to the current viewpoint of the user. Thus, it is almost impossible to look at the data display from an unfortunate angle. Additionally, employing bars may have another advantage. The visually varying but constant width of the bar may serve as depth cue while the length of the bar can be used to display the differing quantitative values (but also distorted by the depth cue it provides). Ware & Plumlee (2005) recommend to use x and y of the screen coordinates to convey information rather than z, the depth of the 3D environment, what is adhered to when using 2D bar displays on billboards.

Judging the length of bars positioned at different locations in two dimensional space is effectively a judgement of length or ”position [on a] non-aligned scale” (Cleveland and McGill, 1984, p. 532). Cleveland and McGill found this to be the third (length) or second (position on a non-aligned scale) most effective mode of representation for extracting quantitative information from graphs in their experiments (most effective mode is position on an aligned scale). Thus, the application of bars differing in size may also be effective for displaying quantitative spatial information in 3D desktop-based virtual environments. To help the interpretation of data values from the data graphics in the 3D virtual environment, reference frames of a fixed size (reference grid) are employed as they may facilitate the judgement and comparison of values displayed as bar lengths. Reference grids have been shown to facilitate comparison between different panels in other contexts (Cleveland (1994), application of Weber's Law in Baird & Noma (1978)) and is recommended for 3D displays (Shepherd 2008).

In this research either single bars displaying univariate data or bar charts displaying multivariate data are used in the evaluations. Bar charts can be regarded as combinations of single bars with single bars stacked and/or arranged beside each other to display several data values at once (table 2.3 for examples of data graphics displaying bar charts).

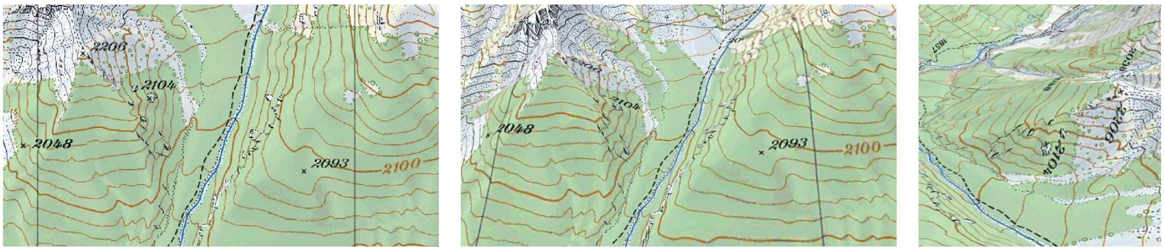
2.3.2. Preparation of 2D and 3D visualisations

For the efficient generation of the above discussed bar and bar chart displays in virtual environments an XML (Extensible Markup Language, W3C 2010*a*) based framework was developed (Burkhard 2008, Bleisch, Burkhard & Nebiker 2009). It allows different data displays with bars as base elements to be created. The bar chart definition is done in a specific XML format defined by a subset of the diagram description XML Schema (W3C 2010*d*) developed by Schnabel (2007) in his thesis about different diagram signatures. The geodata, point data with associated attribute values, needs to be available in a simple XML format defined by an XML Schema (Burkhard 2008). These two definitions are then transformed by a modular XSL (Extensible Stylesheet Language, W3C 2010*c*) transformation and the selected output format generated. For this work KML (OGC 2010) files for the integration of the data displays as placemarks in Google Earth are created. As Google Earth does not support SVG (Scalable Vector Graphics, W3C 2009) placemarks the SVG files are rasterised with the Batik SVG Rasterizer (Apache 2010) to PNG (W3C 2010*b*) files. The whole process is described in detail in (Bleisch, Burkhard & Nebiker 2009, Burkhard 2008).

For the preparation of the 2D representations of the same data sets the process described above for creating KML and SVG/PNG files for placemarks in Google Earth is slightly modified to display the diagrams of bars and bar charts as interactive 2D SVG representation.

The background maps (swisstopo 2010) and ortho imagery used have the same resolution in 2D and 3D (figure 2.6). However, dependent on the viewpoint in 3D not all height labels (spot height and contour lines labels) in the maps or details on the ortho imagery may be similarly readable. Place name labels were removed from the maps to make the mapped areas somewhat anonymous but also to reduce readability problems of north-oriented labels in a virtual environment which potentially can be viewed from different viewpoints. Labelling in virtual environments would need to fulfil special requirements that are not researched in this study but elsewhere (e.g. Maas & Döllner 2006, Maas et al. 2007).

In the remainder of this report the label ’2D’ is used to denote any visualisation in 2D (interactive SVG displays if not otherwise specified) and the label ’3D’ for any visualisation based on 3D virtual environments (based on Google Earth if not otherwise specified).



(a) Map in the 2D SVG representation (b) Map in the 3D virtual environment (c) Rotated 3D

(d) Ortho imagery in the 2D SVG representation (e) Ortho imagery in the 3D virtual environment (f) Rotated 3D

Figure 2.6.: Exemplary background map (swisstopo 2010) and ortho imagery readability

2.4. Research participants

In geovisualization there are typically a small number of geovisualization and/or domain experts working with large data sets in applied settings. To test different geovisualization methods it seems preferable to work with a large group of experts including their data, task and context knowledge. In praxis this is often not applicable. However, data experts from different applied real world settings can valuably participate within selected case studies. For larger controlled experiments or experimental settings, from which statistical significance is expected and to some degree generalisations to a larger populations shall be possible, ’informed’ or ’typical’ participants are needed. In the context of this study, ’informed’ or ’typical’ means that those participants have an interest in making sense of geographically varying quantitative values, have some experience in using (3D) visualisations and thus have skills and abilities in spatial reasoning and navigation. Sebrechts et al. (1999) suggest not evaluating 3D visualisations with short term novice users. This is avoided by using informed participants. Most participants in this study (except the data experts in the case studies) are final semester Geomatics students, former Master of Science in Geographic Information students and staff from different Universities teaching and researching in ’geospatial’ topic areas. This is a somewhat inhomogeneous group but these are the persons most likely to come into contact with the types of geovisualization tested here and thus have some interest in the testing. For that reason, the students are not only an easy accessible replacement for rarely available geovisualization and/or data experts (Marsh 2007) but a valuable and large group of informed users. Nevertheless, the results of the research will bear the characteristics of the sample group and thus may not without further testing be generalisable to a bigger population of potential users of the tested visualisation types.

No participant took part in more than one stage of the research (details of the different research stages in section 2.5). As the research required several years to complete a number of different final semester Geomatics classes could be asked to participate.

2.5. Research methods and implementation

The justification of the research methods (section 2.1) and an overview of the different research stages (table 2.1) are followed by detailed information about task definitions (section 2.2), data graphics design and implementation (section 2.3) and the informed participants of this research (section 2.4). The following sections give details about the methods, tasks, data sets, implementation and procedures, and collected data of each of the four research stages I, IIa, IIb and III. At the beginning of each section the research questions and main hypotheses for the respective stage are given to help with the identification of the goals of each stage. Important process documents and larger size screenshots are available in the appendices and referred to in the respective sections below.

2.5.1. Stage I

Research question: Are methods of quantitative data representation derived from literature about 2D representation methods appropriate in desktop-based virtual environments in an experimental context?

Hypothesis: Typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative values in virtual environments when using appropriate representation methods.

2.5.1.1. Method

The theory constructed in section 2.3.1, that 2D bar displays are an appropriate method for the representation of quantitative data within virtual environments, needs testing. The desired outcome is a measure of how accurately users are able to read data values from such displays even though the bar lengths are visually distorted by depth cues. This is empirically tested in a controlled experiment where the results from 3D virtual environments and from traditional static 2D quantitative graphics can be evaluated for efficiency (task completion time) and accuracy (interpretation error) and compared. Additionally, bars with a reference frame are compared to bars which are frame free (frames may help the interpretation of bar length, section 2.3.1).

2.5.1.2. Tasks

The intention is to find out if typical users are able to identify the differences in the lengths of two bars displaying different quantitative values in a virtual environment. Additionally, the results shall show how accurately such differences can be detected. Thus, the participants have to identify the taller bar in each setting and compare the two bars (estimating how tall the shorter bar is compared to the taller). Additionally, the task of estimating absolute bar lengths from bar displays with frames in the 2D and 3D setting was completed. These are two elementary tasks 'inverse lookup' (estimation of exact bar length) and 'direct comparison' (identifying the taller bar and comparing the two bars) as defined by Andrienko & Andrienko (2006) (section 2.2).

2.5.1.3. Data set

The experiment uses 20 different combinations of two values randomly selected from the range 1–99. The 20 combinations (subsequently named C1–C20) represent the range of possible bar combinations with the smaller bar varying from being between 15% of the length of the taller bar up to 97%. These 20 combinations are displayed at random locations in four different settings: the bars with frames and without frames on a surface in a 3D desktop virtual environment and as static 2D bar charts (figures 2.7, 2.8 and 2.9). The surface used in the 3D setting consists of an undulating part of the real world, made unrecognisable. (Bleisch et al. 2008)

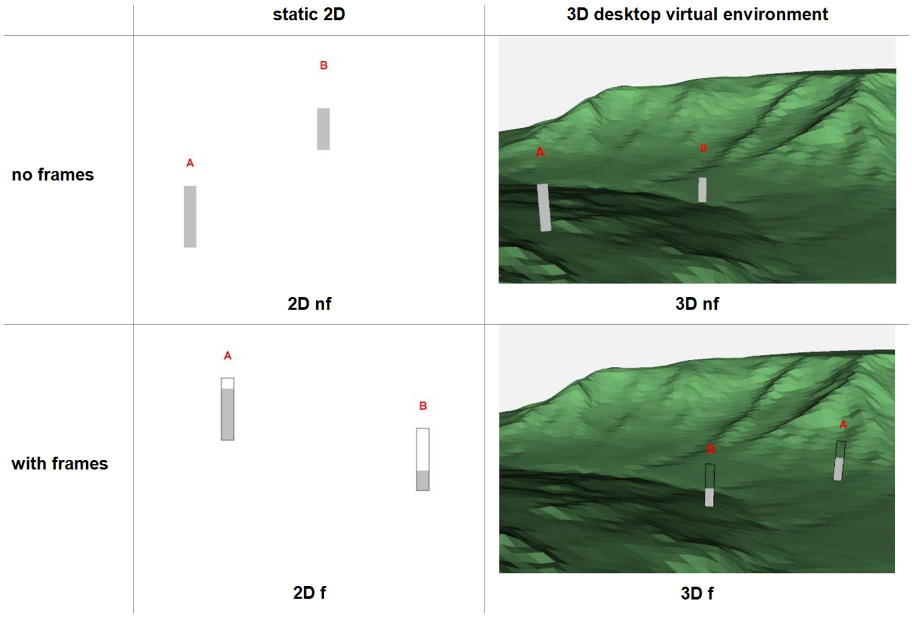


Figure 2.7.: Four different settings of the 2x2 factorial design (static 2D representations without '2D nf' and with frames '2D f' and interactive 3D desktop virtual environments containing bars without '3D nf' and with frames '3D f'), figures 2.8 and 2.9 for all bar combinations C1–C20 (Bleisch et al. 2008)

2.5.1.4. Implementation

The experiment was conducted using a 2x2 factorial within-subject design randomising the order of the experiment assignments (Elmes, Kantowitz & Roediger 2006). The two independent variables are nature of representation (static 2D vs. interactive 3D desktop virtual environment) and ancillary data graphics (bars with and without frames). Whilst the dimensionality of display is not isolated from other aspects of the 3D virtual environment representation (including interactivity and different information content), this representation is referred to as '3D' as the users found this a useful way of describing the distinction. (Bleisch et al. 2008)

26 final semester Bachelor Geomatics students at FHNW University of Applied Sciences Northwestern Switzerland who have some experience of using 3D virtual environments participated in

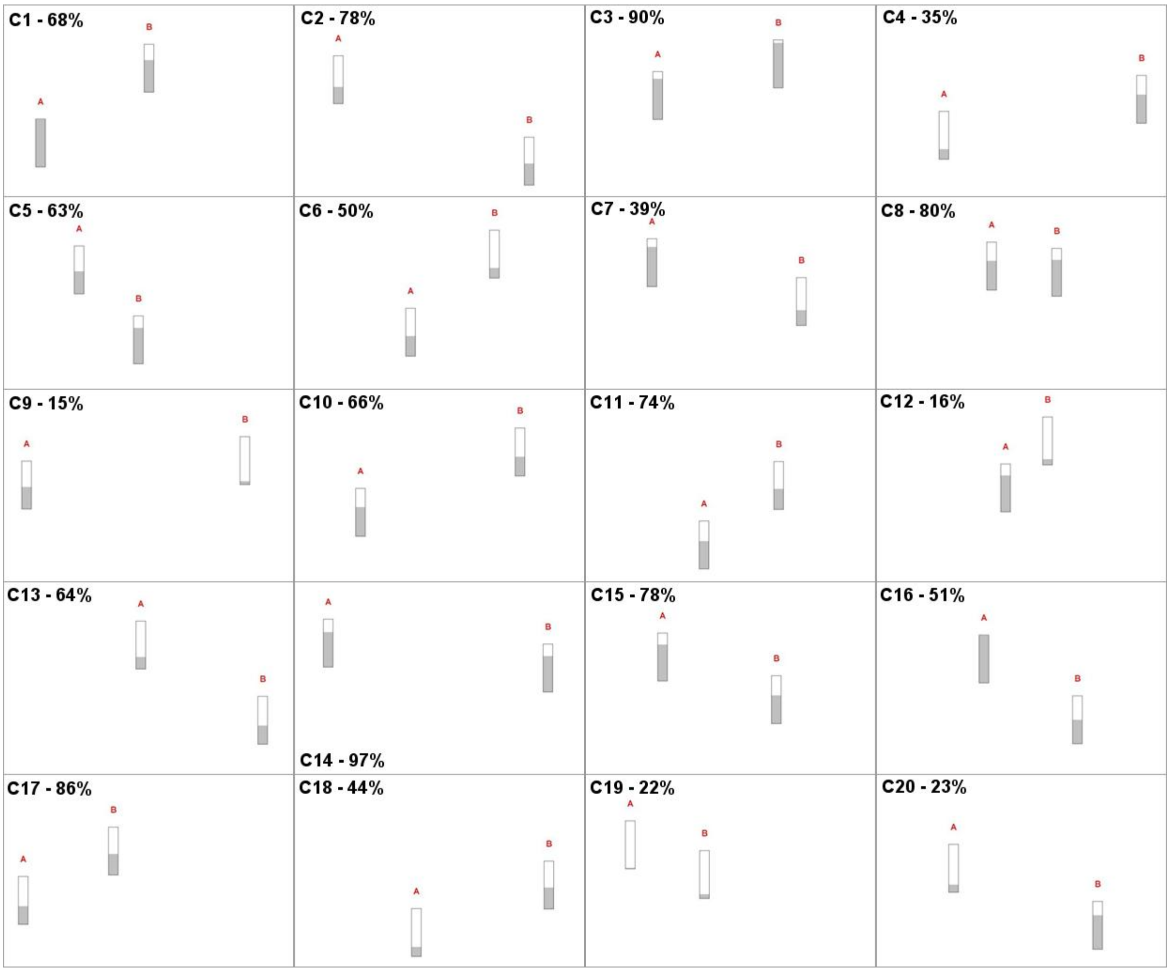


Figure 2.8.: All 20 bar combinations in the '2D f' setting. The proportion of the smaller bar to the taller bar is given as a percentage value ('2D nf' setting: same bar positions but without frames) (Bleisch et al. 2008)

the experiment (section 2.4). Each participant completed each task with a random selection of 10 of the 20 bar combinations in each setting ('2D f', '2D nf', '3D f' and '3D nf'). A subgroup of 18 participants conducted the additional experiment of estimating absolute bar lengths in the '2D f' and '3D f' settings. The information was displayed using JPEG (JPEG 2010) images (2D) and X3D (W3D 2010) environments (3D). The 3D environments were viewed and navigated using the Flux Player software (mediamachines 2010). The participants were encouraged to navigate in the virtual 3D environment if they thought it helped accomplish the tasks but were reminded that this was an efficiency test and that completion times were being recorded. The experiments in 2D and 3D were administered and the participants' answers and task times were recorded using the quiz facility of the WebCT e-Learning platform (now Blackboard, Blackboard 2010) with which participants were familiar. They performed the experiment on generally available desktop computers at the FHNW Institute of Geomatics Engineering in controlled and consistent conditions with the researcher present. After performing the different tasks, the participants had the opportunity to comment in writing on any aspect of the experiment and their performance by answering the last question of the quiz. (Bleisch et al. 2008)

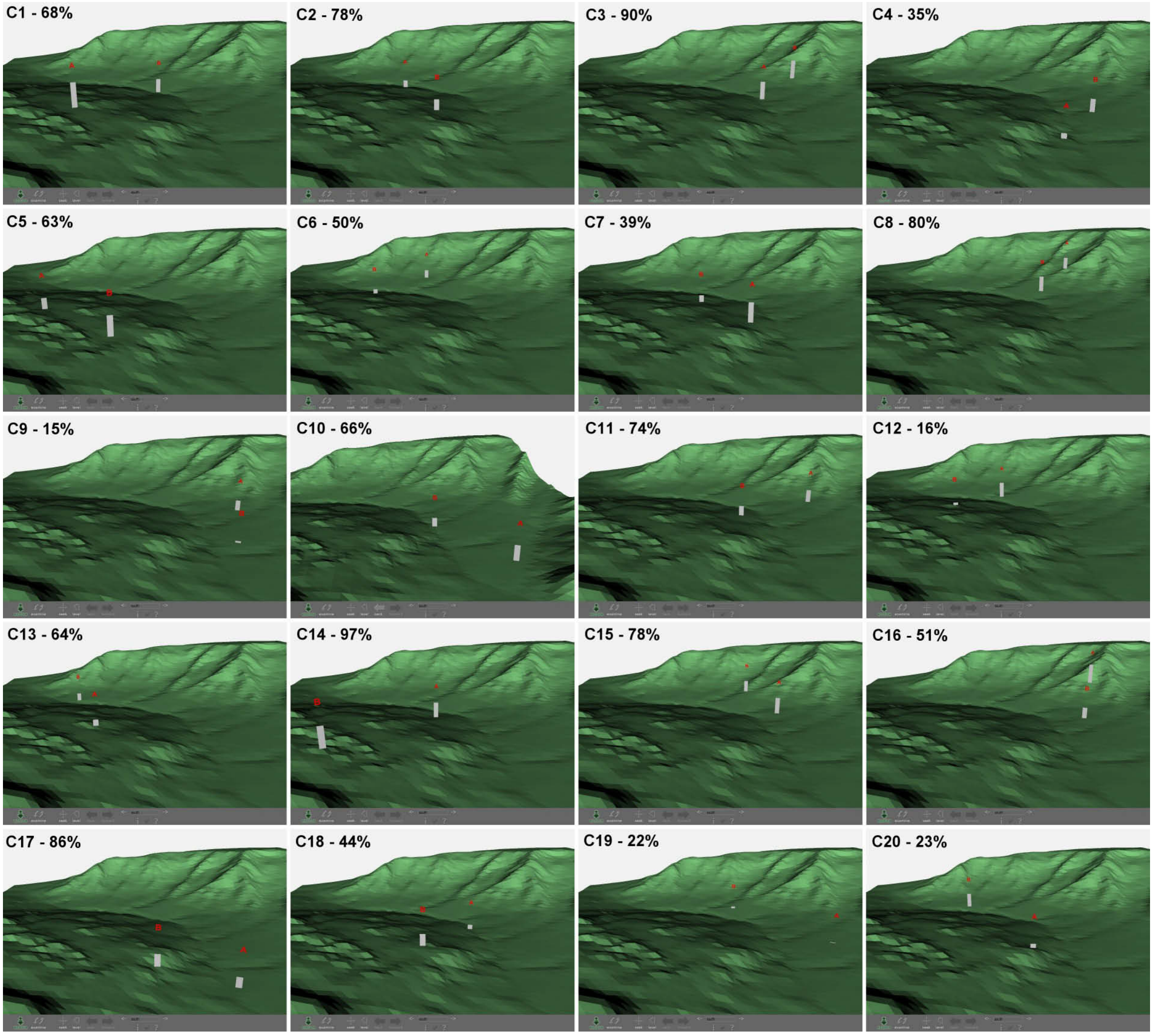


Figure 2.9.: All 20 bar combinations in the '3D nf' setting (screenshots of interactive 3D visualisations from a fixed viewpoint). The proportion of the smaller bar to the taller is given as a percentage value ('3D f' setting: same bar positions but with frames) (Bleisch et al. 2008)

2.5.1.5. Collected data

The data collected for each task included a statement regarding which of the two bars participants judged to be taller (A, B or equal), the percentage value when estimating the length of the shorter bar in comparison to the taller bar and the time needed to fulfil those two tasks (judging the displays and recording the answers). For the 3D settings the time recorded also includes the duration of starting and closing the 3D scene, which takes an average of 3s but depends on the load on the Internet connection. (Bleisch et al. 2008) Additionally, the participant's comments regarding the experiment and their experiences were recorded.

2.5.2. Stage IIa

Research question: Are these methods [2D bar displays] appropriate for more complex visualisation tasks involving landform?

Hypotheses:
Typical users are able to relate multivariate data representation displayed in virtual environments.
Typical users are able to relate appropriate representations of quantitative data to landform.

2.5.2.1. Method

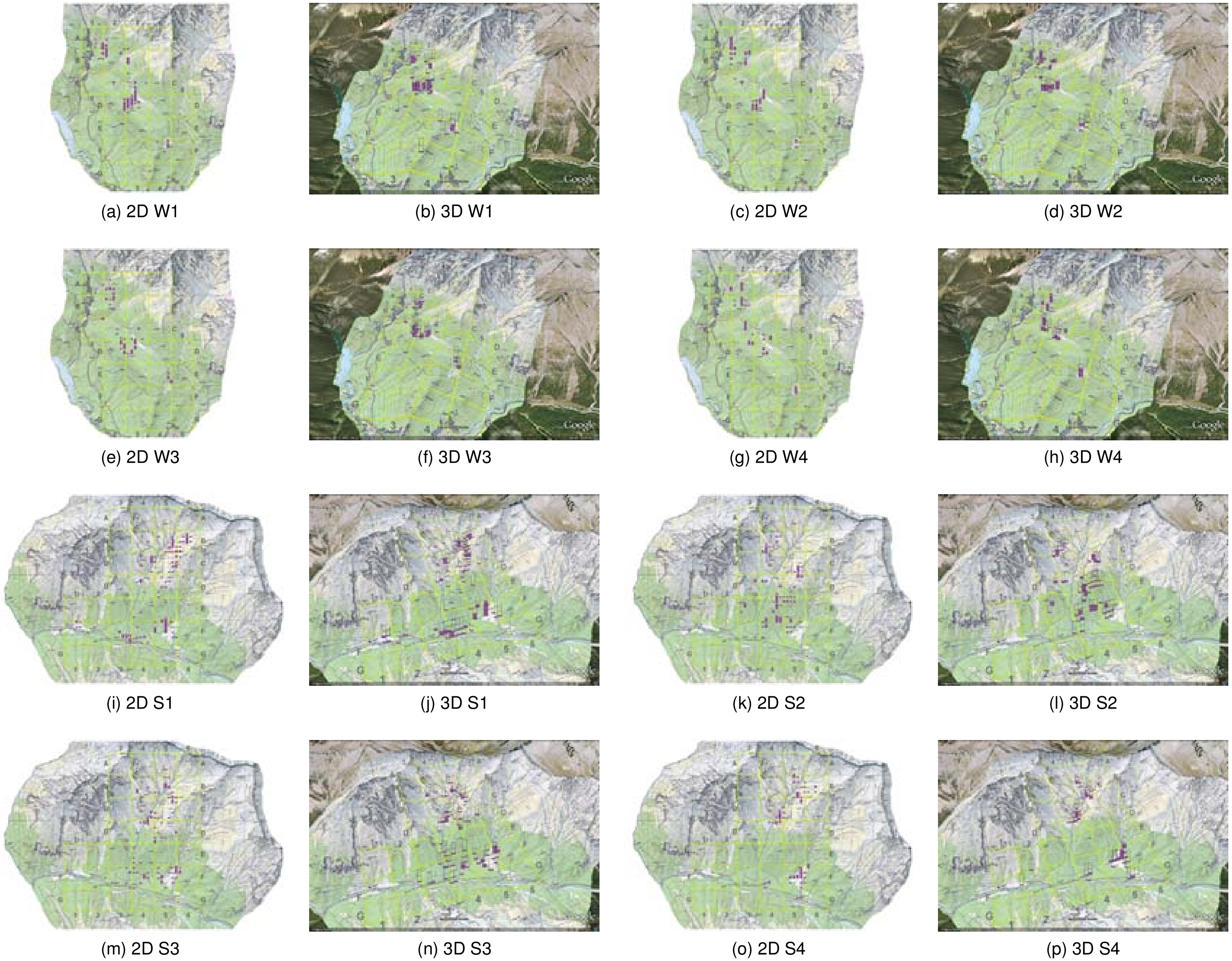
The 2D bar displays shall be tested with more complex data (several bars), more complex tasks (relation of data and landscape) and more context. More context is given by informing the participants what the data and settings are about. Thus allowing participants to potentially include their (probably limited) knowledge of such data sets and settings. The evaluation is done in an experimental setting employing online qualitative questionnaires with a series of tasks collecting open-ended answers, a confidence rating for each answer and the task completion time.

2.5.2.2. Tasks

To test the hypotheses posed for stage IIa a set of tasks which are more complex than bar length estimation and comparison as employed in stage I is required. Stage IIa tasks have to take into account the higher complexity of the data sets and support the relation of data and landscape. Geodata are inherently structured on the base of their location (x and y) and altitude (z) and often unstructured among their non-spatial variables and thus spatial relations have a real meaning in geographic space (MacEachren & Kraak 2001). In addition to the analysis of the data values of a data set per se, the tasks employed in this experimental setting allow analysis of the relations of the data in and to the two- and three-dimensional space as represented by the data value positions (x,y,z) and the landscape they relate to. This is helped by the functional data and task view defined by Andrienko & Andrienko (2006) (section 2.2). It allows the complexity level and reference set for each task to be specified. The tasks can be defined with varying complexity (elementary and synoptic tasks) and different reference sets. Thus, tasks may either refer to location (x,y), altitude (z) or a combination of location and altitude (x,y and z). The exact wording of the tasks and also the writing of the initial instructions for the experiment is defined by following the TAP paradigm (topic, applicability and perspective, Foddy 1994) while trying to keep the tasks as short as possible. Table 2.2 shows the seven tasks (t1-t7) employed in stage IIa. Two of them refer to location (L ref), three of them refer to altitude (A ref) and two refer to location and altitude (LA ref). On the basis of these task definitions the answers can be evaluated regarding their information content in relation to location, altitude or both location and altitude.

task no	task	reference(s)
t1	Which area is most often visited by the deer?	location (L ref)
t2	Which altitude/altitude range is most often visited by the deer?	altitude (A ref)
t3	Generally, compare the number and distribution of deer visits in the lower areas to the number of deer visits in higher areas.	altitude (A ref)
t4	Compare the location and altitude of areas with similar deer visit patterns.	location and altitude (LA ref)
t5	In which area(s) does the number of deer visits increase with increasing altitude?	location and altitude (LA ref)
t6	Describe the deer visit patterns in relation to altitude and landform (surface / topographic features and undulations).	altitude (A ref)
t7	Describe the deer visit patterns in relation to location and land-cover.	location (L ref)

Table 2.2.: Tasks, task numbers and task references as used in stage IIa



(a) 2D W1 (b) 3D W1 (c) 2D W2 (d) 3D W2

(e) 2D W3 (f) 3D W3 (g) 2D W4 (h) 3D W4

(i) 2D S1 (j) 3D S1 (k) 2D S2 (l) 3D S2

(m) 2D S3 (n) 3D S3 (o) 2D S4 (p) 3D S4

Figure 2.10.: 2D and 3D representations of the different data sets (W1-W4 and S1-S4) in the two different areas W (top half) and S (lower half); larger figures in appendix B.1

2.5.2.3. Data set

The data set used in stage IIa was originally collected in the mountainous environment of the Swiss National Park by GPS-tracking red deer (SNP 2010). The data set is composed of coordinates (Swiss Grid LV03) and height values of the places where the red deer stayed or wandered through with a ½ hour resolution. To be able to represent the point data set as bar charts in a virtual environment the locations were transformed to WGS84 coordinates and aggregated in a 0.001° x 0.001° latitude/longitude raster (WGS84). This was done for three months of deer tracking data during winter time and three months during summer time. The location of the winter time data was transformed to an area nearby where other deer live but which has a slightly different topography to have two different study areas (settings W for winter and S for summer) for comparison and evaluation of the influence of different topographies on the data analysis. The data in both areas is from the same red deer but the transformed winter time data has fewer data points and is less spread out than the summer time data. In transforming the winter data to another area it was carefully tried to create a realistic setting which shows some relation between data set and environment. Informally, a data expert was fooled into believing that the deer data belonged to the area it was transformed to. The affine transformation (translation x=-0.0105°, y=-0.0075, rotation r=-1.2°) of the winter data took place before aggregating the point data. In addition to the transformation of the winter data set and aggregation of both data sets, the data was subdivided for different times of the day. Four of these six three-hour data sets were used for the four summer (named S1-S4) and four winter (named W1-W4) data sets. This eight different data sets were then visualised as bar charts in 2D (interactive SVG representation) and in 3D virtual environments (KML placemarks in Google Earth). The background consists of the LK25 map (a very detailed topographic map of Switzerland, scale 1:25'000 (swisstopo 2010)) and an overlaid grid to help with the localisation of the findings. Figure 2.10 shows screenshots of all the data sets in 2D and 3D. See section 2.3.2 for details about the data display creation. Compared to stage I the background is no longer plain topography but topography overlaid with a map. This supports answering more complex tasks with more context information. Employing an interactive 2D visualisation reduces the gap in task completion time between 2D and 3D in stage I but is also a necessity as the data is spread over a larger area and the participants may want to zoom in and out to answer the tasks and read the background map.

According to the findings of stage I (Bleisch et al. 2008) that the tasks are faster completed in virtual environment if the bars have a reference frame, reference frames are also used to test the denser data sets in the virtual environment in the stages IIa and IIb. Single bars without reference frames are used in the 2D representation to avoid the use of display space for frames without additional benefit.

2.5.2.4. Implementation

The data sets (eight in total, each as 2D and 3D visualisation) were administered as online questionnaires (based on XQuest (Hübsch 2008)). Each of the eight questionnaires contains an introduction, a test 2D and 3D visualisation with interaction instructions for 2D and 3D as well as test questions to familiarise the participants with the type of questions asked. Additionally, each questionnaire contains the seven tasks for each of the two 2D and two 3D data sets per questionnaire (two winter and two summer data sets). The different settings and tasks were arranged in a balanced order in the different questionnaires and the questionnaires assigned randomly to the participants (appendix C.1). Study participants were final semester Geomatics students at FHNW University of Applied Sciences

Northwestern Switzerland, staff members from different Universities (University of Zürich, FHNW, City University London) and alumni of the MSc Geographic Information course at City University London. In total 34 of the 74 invited participants took part. Participating in the study and filling in the questionnaire requires at least a computer and software capable of running Google Earth, interactive SVG displays and a fast internet connection. The exact computer configurations cannot be controlled for. We received feedback from 14 potential participants that they were not able to run one or more of the above detailed components (most problems occurred with the operating systems Linux and MacOS). Rather more feedback (from 26 potential participants) contained the information that participation was intended but they could not find the time to complete the questionnaire. The questionnaire took approximately one hour to complete.

Two pre-tests were conducted to test the questionnaires and the completeness and comprehensibility of the instructions within it using the think-aloud method (Nielson, Clemmensen & Yssing 2002) and asking the participants about their experiences after finishing. Short sessions practising 'thinking-aloud' were done with participants who have not used this method before (Boren & Ramey 2000). The questionnaire pre-tests resulted in minimal changes (correction of grammatical errors and slight adaptation/extension of the written instructions). The tasks were reported to be understandable but some of them difficult to fulfil. In addition to the pre-tests, think-aloud studies with five participants (staff members from FHNW and City University London) were conducted. The data recorded from these studies are used to validate and triangulate the findings from the questionnaires. With the think-aloud data it is controlled if the written answers are in accordance with the process of thinking about the data or if the information content is generally smaller or less detailed in the written answers. Comparing the think-aloud protocols with the written answers for each task shows that the final conclusion/insight about the data regarding a specific task is recorded in detail in the written answer. The think-aloud protocols for a specific task generally contain more information than the written answer. However, this information is most often either not relevant, such as first thinking in a wrong direction or misunderstandings which are corrected during the process of answering the task. If some of this 'pre-thinking' is relevant for the final answer then it is also included in the written answer.

2.5.2.5. Collected data

Through the questionnaire the answers to the seven tasks in the eight data sets (four per participant 2x 2D and 2x 3D), a confidence rating (three-step scale: high, medium, low (Rester et al. 2007)) for each answer and the time needed to fulfil each of the tasks are recorded in a MySQL database (Oracle 2010). The time recording includes the thinking about the task and the writing plus the often minimal time needed to decide about the confidence rating and tick the appropriate box before going to the next task. Analysis of the data collected through the think-aloud method showed that it would be difficult to separate the thinking about a task from writing the answer. For all think-aloud participants the process of thinking, writing, checking back with the visualisation was very intertwined and separation of these tasks for the sake of more precise time measurements would almost certainly have meant a loss in richness and detail in the answers. The questionnaires were concluded with the possibility to write about perceived advantages and disadvantages of the visualisation types tested and/or any problems encountered.

2.5.3. Stage IIb

Research question: Are these methods [2D bar chart displays] appropriate for more complex visualisation tasks involving landform?

Hypotheses:
Typical users are able to relate multivariate data representation displayed in virtual environments.
Typical users are able to relate appropriate representations of quantitative data to landform.

2.5.3.1. Method

The 2D bar displays shall be tested with multivariate data (displayed as bar charts). The data set and context information is the same as in stage IIa. The task complexity is the same as the one of the more complex tasks of stage IIa. The evaluation is done in an experimental setting employing insight reporting (Rester et al. 2007). All reported insights are rated for confidence by the participants and collected together with the insight finding time.

2.5.3.2. Tasks

In comparison to stage IIa only one single, explorative task is used in stage IIb to stimulate data exploration and insight generation. One single task is used because the collected data for each task in stage IIa is strongly dependent on the seven different tasks and some of the participants of stage IIa commented that the tasks were difficult to fulfil (section 4.1.4). Additionally, Marsh (2007) recommends using directed exploratory tasks for the facilitation of ideation. The single task for stage IIb is:

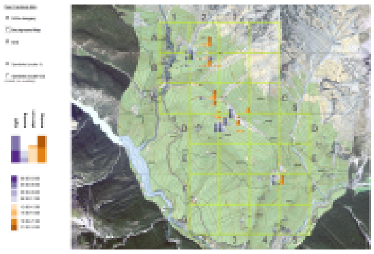
Analyse the deer data and describe the deer habits regarding location, altitude and time of the day.

The task refers to location and altitude as reference set and time as data characteristic and thus allows evaluating the reported insights in relation to location and altitude. It was formulated using a minimum number of words and trying to balance the wording to give each aspect (the references location and altitude but also the variable time of the day) the same weight. Basically, it is an instruction to explore the data and report the insights until the participants think that they have learned everything from the data (North 2006).

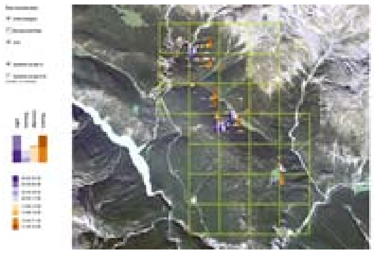
2.5.3.3. Data set

In stage IIb the same data set as in IIa was used. The deer tracking point data was also aggregated in a 0.001° x 0.001° latitude/longitude raster (WGS84) for three months during winter and three months during summer time. The location of the winter time data was transformed to an area nearby to have two different study areas W for winter and S for summer (section 2.5.2.3 for details of the transformation). However, instead of displaying the different hours of the day as different data sets as in stage IIa, all the data was displayed at once. Thus, bar charts were employed to show the length of stay of the deer in an area at different times of the day. The two different data sets for winter and summer were then visualised as bar charts in 2D (interactive SVG representation) and in 3D virtual environments (KML placemarks in Google Earth). The background consists of the LK25 map (a very detailed topographic map of Switzerland, scale 1:25'000 (swisstopo 2010)), ortho imagery with an

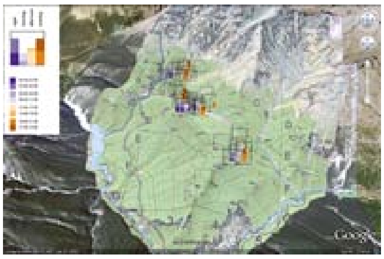
0.5m resolution and an overlaid grid to help with the localisation of the findings. See section 2.3.2 for details about the data display creation. In comparison to stage IIa the ortho imagery was added as background information to the 2D representation similar to the 3D virtual environment of Google Earth. The analysis of stage IIa comments (section 3.6.3) showed that some participants switched on/off the background map to see the ortho imagery provided by Google Earth and that they thought this helpful for the data analysis. Figure 2.11 shows screenshots of all the data sets in 2D and 3D.



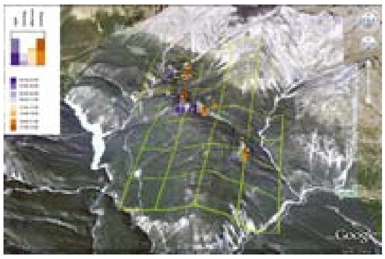
(a) 2D W with background map



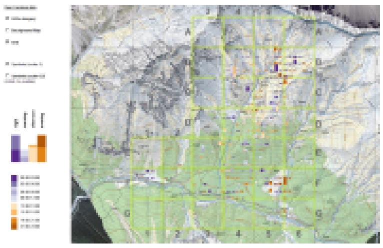
(b) 2D W with background ortho imagery



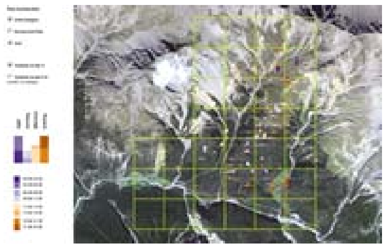
(c) 3D W with background map



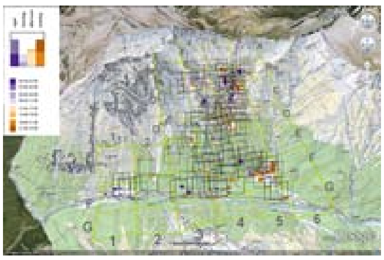
(d) 3D W with background ortho imagery



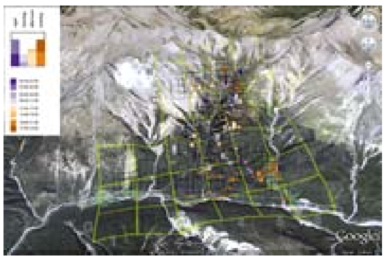
(e) 2D S with background map



(f) 2D S with background ortho imagery



(g) 3D S with background map



(h) 3D S with background ortho imagery

Figure 2.11.: 2D and 3D representations of the two data sets in the two different areas W and S (one data set per area); larger figures in appendix B.3

2.5.3.4. Implementation

The data sets and the task, including several fields for reporting the insights separately, confidence ratings and a comments field were administered as a digital questionnaire (based on XQuest (Hübsch 2008)). Explanations, a test setting and interaction instructions for 2D and 3D were given verbally (appendix C.2). The four data sets were assigned to the participants in a balanced way.

Each participant worked for about half an hour with each of one 2D and one 3D data sets (one winter and one summer data set). The experiment was conducted with 38 final semester Geomatics students at FHNW University of Applied Sciences Northwestern Switzerland which participated in four groups in a face-to-face laboratory setting. This was changed compared to stage IIa because participants in stage IIa spent a lot of time with training and computer problems and many reported that they would have liked to have more support for that. The face-to-face laboratory setting ensured that almost all participants who wanted to participate could do so. Nonetheless, there were three computer problems with loading the virtual environment which could not be solved. All the laboratory computers are supposedly set up the same and potentially capable of running Google Earth, interactive SVG displays and have a fast internet connection. Additionally, participants could ask questions if they did not understand an aspect of the experiment. However, only a few questions needed answered (section 3.6.2).

2.5.3.5. Collected data

The participants recorded their insights into the two different data sets with the help of a number of text fields in a digital questionnaire. The insights recorded (from three to 18 per participant and setting), the time needed to find and write about each insight and a confidence rating (on a three-step scale: high, medium, low (Rester et al. 2007)) for each insight are recorded in a MySQL database (Oracle 2010). The insight reporting is concluded with a last question where participants could comment about experiences and/or problems with the used visualisation type.

2.5.4. Stage III

Research question: Are these methods [2D bar chart displays] appropriate in applied scenarios with real tasks and data experts for the visualisation of quantitative data in relation to the landform?

Hypothesis: Data and task experts gain insights into their data sets when appropriately displayed in virtual environments depicting the landscape to which the data relates.

2.5.4.1. Method

Stage III of the bridge evaluates the 2D bar chart displays within virtual environments in real world applied settings with data and task experts. To study settings within a real-life context, case studies are the preferred strategy (Yin 2003). Gerring & McDermott (2007, p. 688) explain that "the case study is a form of analysis where one or a few units are studied intensively with an aim to elucidate features of a broader class of - presumably similar but not identical - units." Using a case study method, allows testing and further exploration of the findings of the experimental settings in real world settings. This is done descriptively and exploratively (Yin 2003) to understand the characteristics of the real world settings and to find issues that were not addressed by the experimental settings. The explorative nature of this stage is also reflected in the formulation of propositions and research questions instead of hypotheses (table 1.1). Discovered issues may point to further research, potentially also asking for other types of research, for example, for multi-dimensional in-depth long-term case studies (MILCs) (Shneiderman & Plaisant 2006). However, the different stages of this research may be able to answer some issues or questions which arise in the case studies of stage III through integrative data analysis along the research stages and/or by revisiting the data and analysis of previous stages.

To enhance the representativeness of the studied cases a multiple case sampling strategy (Miles & Huberman 1994) selecting diverse cases (Seawright & Gerring 2008) is employed. Studying multiple cases allows for cross-case analysis (Miles & Huberman 1994) which deepens understanding and explanation and adds confidence to the findings as a replication strategy can be followed (Yin 2003) while examining similarities and differences across cases. Three cases, subsequently called case 'Brienz', case 'Literature Atlas' and case 'Deer SNP', were selected which vary in their characteristics. They especially vary in their spatial distribution and the dimensionality of the data, the importance of the landscape in relation to the data and the form and dimension of the landscape in which the data is displayed. This reflects "useful variation on the dimensions of theoretical interest" (Seawright & Gerring 2008, p. 296). The characteristics of the case data sets are described in section 2.5.4.3 and table 2.3.

The comparison between 3D virtual environment and 2D representation as it was part of stages I, IIa and IIb is abandoned in stage III. This is for several reasons. First, the previous experiments have shown that there are only insignificant differences between 2D and 3D displays. Additionally, the experts in the real world settings shall concentrate on the appropriateness of displaying their data within virtual environments as most of them already use some sort of 2D visualisation of their data. To forestall the outlook, it is anticipated that in the future 2D and 3D displays will be used in combination to exploit the strengths of each display type. This, however, shall not be tested within the scope of this research.

In addition to the case study, the data experts of stage III were asked to answer the questions of stage I with the data set with frames in the virtual environment (section 2.5.1). The analysis of this data allows the baselining of the data experts and ensures that they also belong to the group of informed users (section 2.4) who participated in the previous stages.

2.5.4.2. Tasks

No specific tasks are set in stage III as the data experts are expected and instructed to report and fulfil their own real world task needed to answer their research questions in the visualisations.

2.5.4.3. Data sets

In stage III real world data obtained from the data experts in the three case studies 'Brienz', 'Literature Atlas' and 'Deer SNP' is visualised in the virtual environment of Google Earth. Following an overview of the case study data sets. Details are available in table 2.3.

Case Brienz : The case 'Brienz' consists of a data set with 65 points where the coordinates were measured every year from 1989 to 2009. Slope stability is derived from the calculated difference vectors in location and height in millimetres. The points are distributed over an area of 1.5 x 2 kilometres. The observed area consists more or less of one mountain slope. See figures 2.12a-2.12d for example screenshots of the 3D visualisations.

Case Literature Atlas : The case 'Literature Atlas' consists of two data sets, one in central Switzerland and another one in northern Germany. The Swiss data set comprises 147 literature references which are georeferenced and categorised. Six of these categories are used for visualisation. The data covers an area with valleys and mountains in central Switzerland about 40 x 60 kilometres in size. The German data sets comprises 115 literature references which are georeferenced and categorised in different categories than the Swiss data set. Three main categories with three to five subcategories are used for visualisation. The data covers a mainly flat area in northern Germany about 80 x 80 kilometres in size. See figures 2.12e-2.12h for example screenshots of the 3D visualisations.

Case Deer SNP : The case 'Deer SNP' consists of two data sets of deer tracking data in the Swiss National Park (tracking data of deer 604 and deer 635). The data sets 635 was used for the stages IIa and IIb of this research. It consists of 1464 georeferenced deer recordings with date and time of the day for the year 2007. It covers a mountainous area in the Swiss National Park about 3 x 5 kilometres in size. The second data set, deer 604, consists of 5048 georeferenced recordings with date and time of the day for the years 2004-2006. It covers a mountainous area in the Swiss National Park and surroundings about 15 x 30 kilometres in size. See figures 2.12i-2.12l for example screenshots of the 3D visualisations.

2.5.4.4. Implementation

Three case and data experts, one per case study from the different application areas, participated in stage III of the study. A case study protocol was defined according to the instructions by Yin (2003) and Miles & Huberman (1994). The case study process consists of the following steps: set the case study schedule with data experts, experts answer stage I questions and fill in a questionnaire about their data set and context, data visualisations in the virtual environment are created, the data expert is visited and interviewed/observed while working with the visualisation and he/she fills in a short suitability questionnaire and finally, the data expert is re-contacted one week after the visit to capture after-thoughts and thanked for his or her participation. For the data collection empty data 'shells' (Yin 2003) were defined. The questions prepared in the data shells were defined in accordance to the research proposition and questions. They are designed to allow description of the data set and case, show the use of the 3D visualisation during the visit and capturing of after-thoughts. Additionally, the data experts rate the suitability of the 3D visualisation based on three questions. Suitability is a concept basically measuring appropriateness for a purpose (Cowie 1989). Thus, the rather subjective suitability rating of the data experts can be compared to the measures of appropriateness used in the other stages of this research (section 2.1.0.2). Details about the process, the case study protocol and the data collection shells are available in appendix C.3.

2.5.4.5. Collected data

The primary data collected for each case study includes three reports and the answers to experiment of stage I (time and bar length comparison). The first of the reports is a short questionnaire about the data set and the analysis of the data so far filled in by the data experts. The second is the report of the semi-structured interview (questions and answers) and the observations during the interview compiled by the author of this study. The third is a short document created from concluding questions asked in writing by email one week after the interview. Additionally, secondary materials such as existing visualisations or reports of the case or case data set(s) is collected.

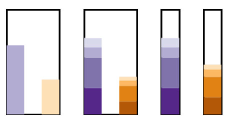
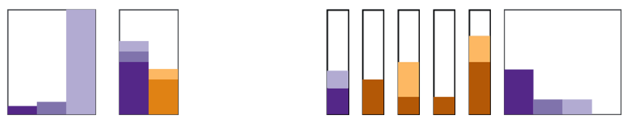
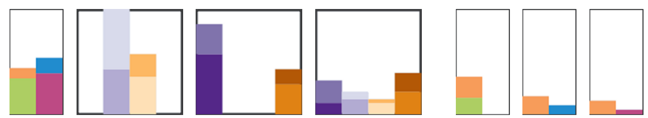
cases / variables	case 'Brienz'	case 'Literature Atlas'		case 'Deer SNP'	
area	1.5 x 2 kilometres Berner Oberland, Switzerland	40 x 60 kilometres Gotthard area in central Switzerland	80 x 80 kilometres Schleswig-Holstein, northern Germany	3 x 5 kilometres Il Fuorn and surroundings, Swiss National Park	15 x 30 kilometres Swiss National Park and surroundings (Zernez to Taufers)
topography	one mountain slope, some local undulations	several valleys and mountains	flat	one main valley and two side-valleys, surrounding mountains	mountainous area
data	65 points, 3D coordinate values (x,y,z) for 20 years (1989 to 2009), position and height difference vectors (mm) calculated between the coordinates of each two subsequent years (one epoch)	147 literature settings (from fictional texts, such as novels, short stories, dramas, tales, ballads, etc. between 1477 and 2005) - categorised (6 categories for visualisation) thematic data with georeference x,y	115 literature settings (from fictional texts between ca. 1750 to the present) - categorised (three main categories with three to five subcategories for visualisation) thematic data with georeference x,y	tracking data of deer 635 - 1464 georeferenced (x,y) recordings of date and time of the day (collected every 7th day during 2007, every 30 minutes for 24 hours)	tracking data of deer 604 - 5048 georeferenced (x,y) recordings of date and time of the day (collected every 7th day for the years 2004-2006, every 30 minutes for 24 hours)
data collection	geodetic measuring equipment (GNSS, tachymetry, levelling), calculation of coordinates (x,y,z)	careful reading and evaluation of the selected fictional texts		two red deers were captured and equipped with a GPS transmitter	
spatial aggregation	none	in 0.001° WGS84 grid	in 0.001° WGS84 grid	in (0.001° and) 0.002° WGS84 grid	in 0.002° WGS84 grid
thematic aggregation	calculation of total difference vectors for five year intervals and for the whole time period (multi epochs)	none	none	aggregation of recordings in three hour intervals (time of the day) for selected months aggregation of recordings per time of the year (col right ↗)	aggregation of recordings per time of the year (Spring [May], Summer [Jun-Oct], Autumn [Nov-Dec] and Winter [Jan-Apr])
created visualisations and their theme (based on data experts input)	- differences in location and height for the total time period - differences in location and height in five year intervals - differences in location in five year intervals - differences in height in five year intervals	- categories 'Tell', 'historisch' and 'freie Fiktion' - categories 'Tell', 'historisch', 'freie Fiktion', 'kulissenhaft', 'aktantenfunktion' and 'unterirdisch'	- categories S and P - cat. S - subcategories 'importiert' and 'transformiert' - cat. S - subcategories 'direkt ref.' and 'indirekt ref.' - cat. P - subcategories 'importiert' and 'transformiert' - cat. P - subcategories 'direkt ref.' and 'indirekt ref.' - cat. P - subcategories 'erinnerter Ort', 'Traumort', 'Sehnsuchtsort' and 'evozierter Ort'	- deer positions per time of the year, year 2007 - deer positions per time of the day, months Jun-Aug 2007 - deer positions per time of the day, months Jun-Aug 2007 (aggregated, 0.001° WGS84) - deer positions per time of the day, months Nov-Feb 2007	- deer positions per time of the year, year 2004 - deer positions per time of the year, year 2005 - deer positions per time of the year, year 2006
example data graphics for each visualisation					

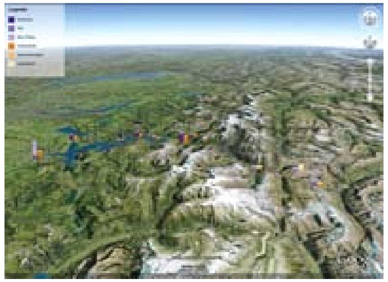
Table 2.3.: Overview of the characteristics of the case study areas, data sets and visualisations


(a) Brienz: fs and dh in total


(b) Brienz: fs per 5 years

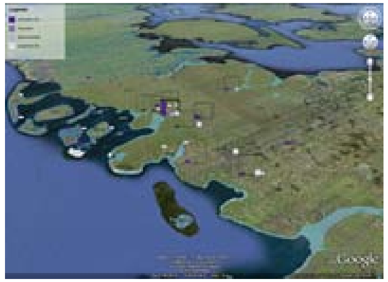

(c) Brienz: dh per 5 years

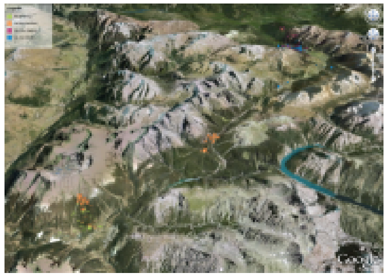

(d) Brienz: fs and dh per 5 years

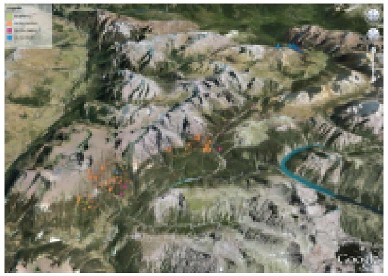

(e) Literature Atlas: area Gotthard

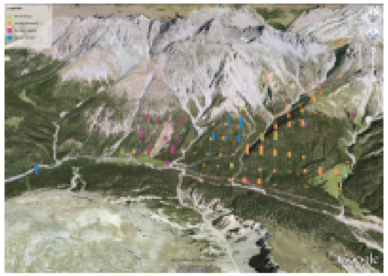

(f) Literature Atlas: detail in Husum NFR


(g) Literature Atlas: area NFR


(h) Literature Atlas: area NFR


(i) SNP: Deer 604 in 2004


(j) SNP: Deer 604 in 2005


(k) SNP: Deer 635 in 2007


(l) SNP: Deer 635 in winter 2007, different times of the day

Figure 2.12.: Exemplary 3D representations of the different case study data sets and areas; larger figures in appendix B.4

2.6. Analysing the data

2.6.1. Quantitative analysis methods

Collected data of quantitative nature, such as time data, bar length comparison data or generally categorical data, is quantitatively analysed using a number of different statistical tests depending on the data types as detailed in the following sections. The decisions for accepting or rejecting the tested hypotheses is always based on a significance level of 95% (α = 0.05). To allow the readers of the analysis chapter 3 their own judgements the test statistic values together with the associated p-values are reported for each statistical test. The statistical analysis of the collected data is done in R (R 2011) and MS Excel Versions 2003 and 2010 (Microsoft 2010*b*).

2.6.1.1. Analysing time data

The times recorded in stages I, IIa and IIb are in seconds. As would be expected for time data of short tasks these data sets are not normally distributed (no negative times, large number of short answer times and some very long answer times). Testing for normality of the data set is done using the Shapiro-Wilk test of normality (R function $shapiro.test(data)$). Comparing two not normally distributed time data sets is done using the non-parametric Wilcoxon Mann-Whitney Rank Sum test (R function $wilcox\_test(data)$ in the coin package).

2.6.1.2. Analysing bar length comparison data

The differences data from comparing two bar lengths in stage I is undertaken with a student t-test (in MS Excel) tested for differences of the mean from 0 which are expected to be normally distributed. Differences between the data sets of the different settings are evaluated using ANOVA and post-hoc Tukey-Tests (Zar 1984) (both in MS Excel, the latter manually).

2.6.1.3. Analysing categorical data

The categorical data, such as the confidence ratings or the results from some of the qualitative analyses (e.g. complexity ratings or word counts, section 2.6.2), is tested for differences using the non-parametric χ^2 (Chi-squared) test (R function $chisq.test(data)$). This test is not reliable if the contingency table contains cell values < 5. The common recommendation (e.g. Papula 2001) to aggregate such cell values does not make sense in cases where the aggregated values are no longer meaningful (e.g. aggregating numbers of insights of low and medium plausibility when plausibility ratings are done on a three step scale only). In the few cases with very small values the test was not done and instead the visualisation of the data is commented upon regarding potential trends.

2.6.2. Qualitative analysis methods

Collected data of qualitative nature, such as task answers, reported insights, think-aloud protocol and transcripts of interviews and observations, is qualitatively analysed by following qualitative data analysis processes mainly guided by Miles & Huberman's (1994) book "Qualitative data analysis:

an expanded sourcebook". The qualitative text data is manually and automatically coded and annotations are made for later detection and linkage of concepts depending on the hypotheses and research questions defined. The following sections give details about the processes and codes. The qualitative data analysis is done in ATLAS.ti (ATLAS.ti 2010).

2.6.2.1. Analysing insights for complexity and plausibility

In stage IIa and IIb qualitative statements, either task answers in stage IIa or so called insights (North 2006) in stage IIb, are recorded. However, the recorded answers and some of the insights resemble short reports rather than single insights into the data (based on the definition of insight by Saraiya, North & Duca 2005). Thus, the recorded texts were split into several basic insights which could be rated. This has the advantage that the insight ratings from stage IIa and IIb can be compared. On the other hand, the time recordings and confidence ratings are per answer (stage IIa) or participant's insight (stage IIb) and are thus somewhat less comparable between the two stages.

The insights of stages IIa and IIb are evaluated for complexity and plausibility. Complexity is rated on a three-step scale: low, medium and high (Rester et al. 2007) depending on the number of values described or analysed or even the inclusion of context information leading to hypotheses. Table 2.4 explains the three complexity levels with explanations and examples. Saraiya et al. (2005) rate the correctness of the insights in their study. Rester et al. (2007) find it difficult to rate correctness and propose classifying the insights on a three-step scale of plausibility (Rester et al. 2007). This is also done in the research presented here. All the insights are rated for low, medium and high plausibility. The low plausibility category includes the few cases of not plausible insights.

complexity level	explanation	examples (insights categorised in the respective complexity level)
low	insights of low complexity, referring to one or only few values or a simple fact	"area F5" "about 1935m" "lowland valley area around grid squares 3F to 5F"
medium	insights of medium complexity, referring to several values, comparing and analysing	"and below 2100 I'd say that lower altitude ranges have more visits and are more clustered" "The patterns with a low number of visits are distributed over the entire surface at an altitude of about 2000m" "Most concentrations on upland valley sides, both NW and SE facing."
high	insights of high complexity, referring to a multitude of data, evaluate, find hypotheses about the data (taking into account context information)	"They would appear to prefer SE-facing slopes (probably warmer)" "it then appears to increase with altitude if the more loftier areas are potentially sheltered and not on a ridge." "I think what I am seeing is visits over a period of time that embraces deer moving from relatively low (2000m) up to the snow line higher up, presumably to exploit vegetation in the warmer months."

Table 2.4.: Explanations and examples of complexity levels for insight rating

Insight reporting often includes domain specialists as coders or for additional input in the analysis (Saraiya et al. 2005, Rester et al. 2007). This is not feasible in this study as real data sets are used but part of the data is transformed to an area nearby to simulate a second different but similar setting for evaluation and comparison (sections 2.5.2.3 and 2.5.3.3). Additionally, the study participants are informed users of the visualisation environments and general analysis of geospatial data but are not data experts. The complexity and plausibility is rated by the author based on the contents of the reported insights and a comparison of the insight contents with the visualisations. North (2006) admits that insight coding is inherently subjective and recommends maintaining objectivity through strict coding practices. This recommendation is followed as described in section 2.6.2.2.

2.6.2.2. Process of manual coding

All insight ratings are carried out as consistently and completely as possible by the author. This is achieved by coding large portions of text at a stretch. In a further step all the text passages coded within a category are compared and outliers reconsidered and potentially assigned to another category. To ensure intra-coder reliability, several parts of the text were coded a second time after at least two weeks and compared to the first. Only minimal insignificant differences between the two codings were detected. Thus, it is concluded that consistency in coding was achieved. The number of insights rated in the different complexity and plausibility categories are finally quantitatively analysed (section 2.6.1.3).

Intercoder reliability (Neuendorf 2002) was tested for by having, in addition to the author, two independent researchers (subsequently labelled I and II) code large parts of the text representing answers to different settings and tasks. These where then compared to the author's (labelled B) original coding results for the respective parts of the text. Researcher I coded 652 insights (92%) within the answers for the settings S2, W1 and W3 as compared to 711 insights by the author. Researcher II coded 316 insights (114%) within the answers in all settings for question t3 as compared to 278 insights by the author. Figures 2.13, 2.14, 2.15 and 2.16 show the complexity and plausibility ratings for the respective settings and tasks between the researchers I and II and the author (B). The complexity and plausibility ratings are tested for significance between researcher I or II and the author in 2D and 3D and within the same coder between 2D and 3D using χ^2 tests. The differences are all insignificant but for the difference between author and researcher I in the 2D complexity ratings (χ^2 = 6.0802, df = 2, p-value = 0.04783) which is just significant. Comparing the codings of the different researchers shows that the largest variability occurs in the decision about what portion of the answer texts constitutes an insight. Splitting the answers into more insights leaves the insights less complex and potentially more plausible while splitting the answers into less insights leaves them more complex and potentially less plausible. This variation is stronger for stage IIa where most of the answers need splitting into insights, and it is less strong for stage IIb where the participant's reported insights and only some longer insight reports need splitting. The difference between 2D and 3D is insignificant for all coders. The variation in complexity and plausibility between the author and researcher I or II varies more than between 2D and 3D but is mostly insignificant (except the one relationship mentioned above). Thus, there are slight differences in the interpretation of the different coding categories but the codings by each researcher are consistent. Nevertheless, this minimal variation which could occur when the whole data set would be analysed by different researchers than the author may be taken into account when considering the results of the research.

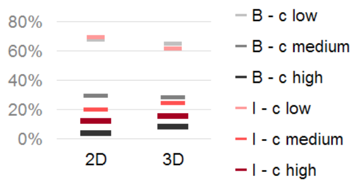


Figure 2.13.: Comparison of the complexity ratings (high, medium, low) of the insights between the author (B) and researcher I in 2D and 3D

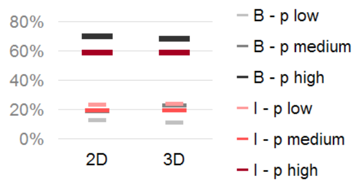


Figure 2.14.: Comparison of the plausibility ratings (high, medium, low) of the insights between the author (B) and researcher I in 2D and 3D

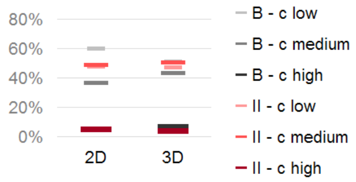


Figure 2.15.: Comparison of the complexity ratings (high, medium, low) of the insights between the author (B) and researcher II in 2D and 3D

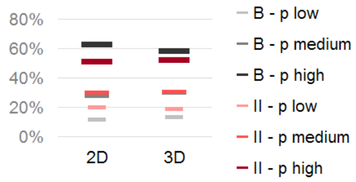


Figure 2.16.: Comparison of the plausibility ratings (high, medium, low) of the insights between the author (B) and researcher I in 2D and 3D

2.6.2.3. Analysing insights for reference type

The insights are in addition to complexity and plausibility also manually coded depending on the reference set (location, altitude or a combination of both) they refer to. This allows the comparison of insights relating to different references between stages IIa and IIb as in stage IIb with only one task set it is not possible to differentiate between tasks with different reference sets. The assigned codes are location (L ref), altitude (A ref) and both (LA ref), the same codes as for the tasks categorisation per task reference in stage IIa (section 2.5.2.2). These codes are based on the coding schema explained in section 2.6.2.4 and table 2.5. Thus, an insight referring to location (L ref) contains either an exact references such as a grid square (e.g. E5) or an indirect reference in form of landcover (e.g. forest, grassy, etc.). Additionally, terms like 'north of' or 'to the left' are coded as location references. Insights coded as referring to altitude (A ref) contain either an exact height, such as 2345 m or a statement regarding the landform (e.g. slope, steep, etc.). Additionally, also words like 'above' or 'below' are coded as altitude reference. Some insights contain words referring to both location and altitude and are thus coded as referencing both (LA ref).

2.6.2.4. Analysing word counts

Another way of analysing the content of the answers of stage IIa and the insights reported in stage IIb is automatically analysing the frequencies of the words used within the participant's writing. This allows, to some degree, triangulating the findings of the manual encoding with the findings of an automatic technique. According to the research questions and hypotheses posed (section 1.4), a priori words relating either to location or to altitude are interesting (deductive approach to coding, Lewins & Silver 2007). A word crunch analysis (ATLAS.ti 2010) yields frequency counts for each word used in the texts. Based on the hypotheses which propose more efficiency when using the 3D visualisation for landform tasks (or the 2D visualisation for location related tasks) the list of word counts is analysed regarding location and altitude words and three subcategories emerged - datum, object and relation/description within both main categories (emergent vs. a priori coding, Stemler 2001). Table 2.5 lists the six categories and subcategories of location and altitude words analysed and coded.

2.6.2.5. Analysing case study data

The data which is collected in the three case studies contains descriptions and explanations of the case data sets, contexts and processes as defined in the case study protocol and the data collection shells (appendix C.3). This mainly textual data is analysed and displayed by employing cross-case

category	subcategory	example words	category short name
location	datum	D3, F4 (grid square)	L1
	object	forest, scree, grassy	L2
	relation/description	north, left, south-east	L3
altitude	datum	1950m, 2400m	A1
	object	mountain, slope, ridge	A2
	relation/description	steep, lower, highest	A3

Table 2.5.: Word categories and respective example words (the category short names are used in the figures in the analysis chapter 3)

'thick matrix displays' (a combination of 'thick' descriptions (Gerring 2007), which include much detail, and matrix displays (Miles & Huberman 1994) which allow structured viewing of the data). This enables understanding each case in its own terms but at the same time allows comparative cross-case analysis. Transferring the collected data, including descriptions, data experts statements and interview notes to the matrix displays the following informal rules are adhered to if possible and sensible in the research context:

1) similar information is reported together
2) repetitions are left out but words may be introduced to emphasise
3) information not relating to either the research propositions or questions nor being important context or descriptive information is left out
4) information may be shortened if the meaning can be retained or reported fully if this is not the case.

These are not general rules for data condensation and reporting. However, Miles & Huberman (1994) recommend reporting on data transformation processes and, additionally, displaying as much data as sensible (section 3.5.2) even though neither of which is common practice in qualitative research but it enhances validity through reproducibility.

Analysis of the case data is done according to the research questions (table 2.1) based on the matrix displays and checking back to the original data if required (what potentially results in an update of the matrix display). The two research propositions are confirmative and are answered using summarising content analysis strategies including aspects of all three cases if applicable. The third explorative research question asks for comparative cross-case analysis. Themes which cut across cases are looked for and a replication strategy (Yin 2003) is followed. Thus, patterns found in one case study are looked for in the other two cases. Additionally, special attention is given to differences in similar patterns and potential explanations. As purposely diverse cases were selected (section 2.5.4.1) the variations in the case characteristics are of special interest in this regard.

Beyond analysing the case data sets in regard to the research propositions and question, the description of the case characteristics, the processes, meanings and explanations of the data experts shall give the reader information they can compare with their own (potential) applications. The condensed and summarised 'thick matrix displays' are integrated as tables (section 3.5.2) in the analysis part of this report. They form part of the results and allow readers comparisons with other or own applications.

2.6.2.6. Analysing participants comments

Participants' comments are used to triangulate the findings from quantitative analysis methods but also to integrate the research stages along the bridge from experimental to case study methods and vice versa. Triangulation is valuable for the research process as it may reveal convergence, inconsistency or contradiction of the findings allowing the researcher to "construct superior explanations" (Denzin 1978 cited in Johnson, Onwuegbuzie & Turner 2007, p. 115). Content analysis of participant's comments follows a quantitative, summarising approach (Neuendorf 2002). Comments are assigned to categories and quantitatively compared (relative to the number of participants in each stage). Meaning and detail is added to the numbers by summarising the comments and giving examples of each category. However, single comments on rarely mentioned aspects may also shed light on a specific issue. Such comments are reported during the analysis and it is stated that it is the expressed view of a single or a few participants.

All participants' comments are taken into account for the analysis independent of the place of their occurrence. Many participants noted comments or commentary statements within the question answers or the insight reporting of stages IIa and IIb. Few comments were sent by email after conducting the experiments. Additionally, the questionnaires of stages IIa and IIb contain special sections where participants were asked to comment about the visualisations (sections 2.5.2.5 and 2.5.3.5).

2.6.2.7. Analysing think-aloud reports

The think-aloud sessions in stage IIa are recorded in writing and compared to the written task answers. It is found that the written answers contain the same information as what the participants thought, or at least what they expressed orally, during the process of answering. Exceptions to this are dead ends of thinking and recognised wrong thinking which occasionally was encountered but not reported in writing. Thus, the think-aloud protocols are not evaluated further and the written task answers given during the process analysed together with the recorded experiment data of the other participants of stage IIa.

2.6.3. Data handling

The use of several tools such as MySQL databases (Oracle 2010), MS Excel (Microsoft 2010*b*) or R (R 2011) allows a mostly digital data flow, for example, via query languages such as SQL (Structured Query Language) or data import into R from MS Excel via .csv files. This minimises typing errors or errors introduced by copying and pasting data. Nevertheless, the collected data of each research stage (sections 2.5.1-2.5.4) has its specific characteristics which are, together with suitable arrangements for their handling, explained below.

2.6.3.1. Stage I

Starting up the 3D scenes took about 3 seconds. This time was subtracted from the recorded task times before analysing it. Three of the 1400 bar length differences data sets collected were ignored because of null responses. The data were checked for errors and inconsistencies, such as participants using 0 instead of 100 to record equality or entries using decimals instead of percentages for comparison. Such cases were corrected before analysis. From the 360 data sets of absolute bar length estimation collected, 24 were ignored as in these cases participants compared the two different bars rather than estimating the absolute lengths of the two bars. (Bleisch et al. 2008)

2.6.3.2. Stage IIa

Time: The questionnaire automatically recorded the exact computer time whenever the participant changed from one task to the next. For the analysis the time in second per answer was calculated from the differences of the two computer times recorded at the start and end of each task. Times from tasks with empty answers where excluded from the evaluation. They resulted either from participants being distracted (which resulted in a longer than normal task answer time) or participants clicking twice and thus skipping a task (which results in a task answer time of very few seconds). The time data from both cases would distort the evaluation. The recorded times are per task answer. The task answers were subdivided into insights for the content analysis. Thus, the recorded times often span more than one insight. Separation is not possible as it is not known how much time was spent thinking about and reporting each single insight. This is an inherent data characteristic which cannot be corrected for. It has to be taken into account when considering time analysis in chapter 3. Appropriate remarks are added to the analysis chapter to guide interpretation. For general comparisons of insights a pro rata splitting of the times per answers are done to get an indication of time spent per insight.

Con dence rating: The confidence rating of the answers are included in the analysis where given. The occasional cases where the participants forgot to rate their confidence are not included. The confidence ratings, similar to the time measurements, are collected per task answer and may thus span more than one insights. This is an inherent data characteristic which cannot be corrected for. It has to be taken in account when considering the analysis of the confidence ratings in chapter 3.

Task answers: Task answers are included into the analysis where given. The few cases of empty answers were not incorporated. For example, one participant only did the first half of the questionnaire and stated that he did not have the time to finish. Thus, the given answers were included in the analysis the second half of empty answers ignored.

From the seven tasks (section 2.5.2.2) three have the reference altitude, two reference location and two a combined reference of location and altitude. The quantitative analysis of, for example, insights per reference type yields thus naturally more insights for reference altitude. This has to be taken into account when reading the analysis chapter 3 for stage IIa data.

Language: Task answers and comments were given in German and English as staff and students from Switzerland and the UK participated. The analysis and integration of the two languages was done with the authors command of English and German. Additionally, it was tried to take into account that most Swiss participants know the symbology of the background map LK25 very well while the rest of the participants might have been seeing it for the first time. Thus, for example, ”green area” and ”forest” were judged as referring to the same type of land-cover and thus rated consistently. A legend for the map symbolisation was not provided.

Comments: Comments within the task answers were identified and evaluated together with all other comments and the notes from the comments section of the questionnaire.

2.6.3.3. Stage IIb

Time: The digital form used for insight reporting recorded the exact computer time whenever the participant changed from reporting one insight to the next. For the analysis the time in second per insight was calculated from the differences of the two computer times recorded at the start and end of each insight. Times from empty insight reporting fields were excluded from the evaluation. They resulted from participants clicking twice and thus skipping a form field (which results in an insight reporting time of very few seconds). Such time values would distort the evaluation. Additionally, some participants wrote more than one insight per field. This cannot be separated in terms of time. Thus, the recorded times sometimes span more than one insight. This is an inherent data characteristic which cannot be corrected for. It has to be taken into account when considering time analysis in chapter 3 (section 2.6.3.2, for remarks on time in stage IIa).

Con dence rating: The confidence rating of the insights are included in the analysis where given. The occasional cases where the participants forgot to rate their confidence are not included. Confidence ratings are per recorded insights and may thus occasionally span more than one evaluated insight as some longer insight reports were recorded which required splitting.

Reported insights: All reported insights are included in the analysis. Empty insights were not taken into account. Rating the insights for complexity, plausibility and references (section 2.6.2) results in a different number of insights in each of these categories as some insights cannot be rated within one or several categories. For example, the insight ”The deer is more active during the night” does not have reference either to altitude or location but it can be rated for complexity and plausibility. This characteristic is inherent to the data and not corrected for. It has to be taken into account when looking through the analysis chapter 3.

Comments: Comments within the task answers were identified and evaluated together with all other comments and the notes from the comments section of the questionnaire.

Language: All questionnaires were completed in German and the results later translated to English. The analysis and integration of the two languages was done with the authors command of German and English.

2.6.3.4. Stage III

Interview recording: The interviews and observations during the visits at the data experts workplaces were recorded in note form by the researcher. The notes were re-read and potentially revised for better understanding the day of the interview or the following day.

Language: All questionnaires for and interviews with the case study data experts were done in German and the results later translated into English where needed. The analysis and integration of the two languages was done with the authors command of German and English.

3. Data analysis and results

< Chapter 2. Methods

This chapter is informatively summarised in the discussion chapter (4). This summary is thus descriptive including recommendations for reading.
The contents of this chapter consist of the description, statistical testing and graphical displays of the collected data sets in the four research stages I (3.1), IIa (3.2), IIb (3.3) and III (3.5). The results are reported but free from discussion and interpretation (Wilkinson 1991). The integration, interpretation and discussion of these results is done in the discussion chapter (4). There, all results are referenced and linked back to this chapter for details. For stage III (3.5), the collected qualitative data is compiled in detailed content rich tables. Additionally, the analysis of the participants' comments (3.6) is running text providing much detail information.
It is recommended to only read the sections containing the compiled data of stage III (3.5) and potentially also the analysis of the participant's comments (3.6). Later, while reading the discussion chapter (4), it may be valuable to refer back to the first three sections of this analysis chapter for details about the results of stage I, IIa and IIb.

Chapter 4. Discussion >

3.1. Stage I - Evaluating single bars with simple tasks

The description and analysis of the data collected in stage I is published in (Bleisch et al. 2008, p. 218ff, ➩ Appendix A). The following sections are extracts of the publication.

3.1.1. Description of the sample

The experiment of stage I was conducted with 26 participants. The data collected for each task included a statement regarding which of the two bars participants judged to be taller (A, B or equal), the percentage value when estimating the length of the shorter bar in comparison to the taller bar and the time needed to fulfil those two tasks (judging the displays and recording the answers). For the 3D settings the time recorded also includes the duration of starting and closing the 3D scene, which takes an average of 3 seconds but depends on the load on the Internet connection. The accuracy of the comparison of the two bars in each setting is evaluated using the difference between the participant's estimated value and the actual percentage value (e.g. the smaller bar is 67% of the taller bar, if the participant's estimation is 70%, then the difference of +3 is used for the evaluation of the judgement accuracy). The expected values for these differences are 0 for precise estimations by the participants. There is a slight bias in the data as people tend to estimate the differences in 5 or 10 s (e.g. estimation of 65 or 70% and not 67%). But the data compensates for this bias as under and over judgements can be assumed to be made with equal frequency. (Bleisch et al. 2008, p. 218f, ➩ Appendix A)

3.1.1.1. Time

The log normally distributed task completion times in seconds in 2D with (2D f) and without reference frames (2D nf) and 3D with (3D f) and without reference frames (3D nf) are illustrated in figure 3.1. Using ANOVA and post-hoc Tukey-Tests to test the hypothesis that the means of the times per setting are equal ($H_0 : m_{2Df} = m_{2Dnf} = m_{3Df} = m_{3Dnf}$) results in rejection of H_0. (Bleisch et al. 2008, p. 221, ➪ Appendix A)

3.1.1.2. Bar length differences

Applying a Student t-test shows that the means of the differences between estimated and actual values for the settings '2D f', '2D nf' and '3D f' are significantly different from 0 (figure 3.2). The mean of the differences in setting '3D nf' is not significantly different from 0. Using ANOVA and post-hoc Tukey-Tests to test the hypothesis that the means of the differences per settings are equal ($H_0 : m_{2Df} = m_{2Dnf} = m_{3Df} = m_{3Dnf}$) results in the rejection of H_0. The alternative hypothesis ($H_1 : m_{2Df} = m_{2Dnf} = m_{3Df} \neq m_{3Dnf}$) is accepted. (Bleisch et al. 2008, p. 220f, ➪ Appendix A)

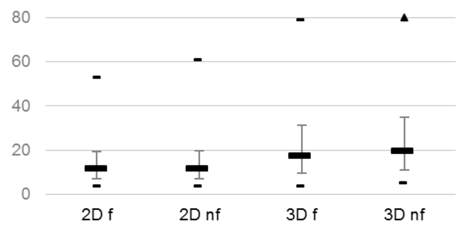


Figure 3.1.: Mean, standard deviation (calculated from log normal distribution), and minimum/maximum values (max value 3Dnf = 98) of task completion time in seconds in all four settings, 3 seconds of 3D scene start-up time were subtracted in the two 3D settings '3D f' and '3D nf' (Bleisch et al. 2008, p. 220, ➪ Appendix A)

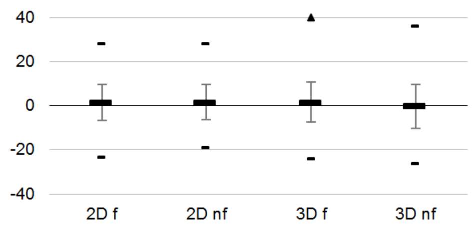


Figure 3.2.: Mean, standard deviation, and minimum/maximum values (max value 3Df = 45) of the differences between estimated and actual values in all four settings (Bleisch et al. 2008, p. 220, ➪ Appendix A)

3.1.1.3. Identi cation of the taller bar

The participants were able to identify the taller bar in all four settings (2D and 3D with and without frames) in almost 100% of the cases (figure 3.3). In each setting, a number of bar combinations, mostly with proportions higher than 80%, were judged as being equal in size (figure 3.4, and shown by the medium grey portions in figure 3.3). (Bleisch et al. 2008, p. 219, ➪ Appendix A)

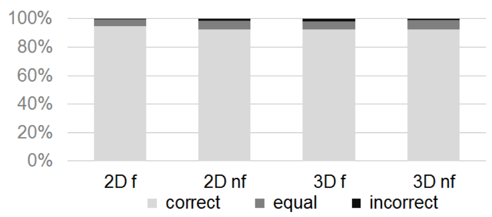


Figure 3.3.: Correct and incorrect judgements of the taller bar. Medium grey values indicate bar combinations judged as being of equal size (figure 3.4) (Bleisch et al. 2008, p. 220, ➪ Appendix A)

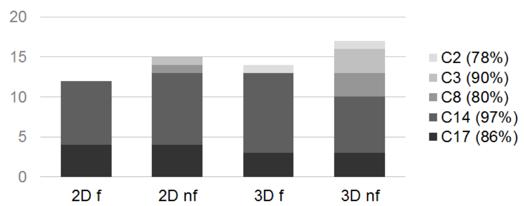


Figure 3.4.: Frequency of bar combinations judged as being of equal size (proportion of the two bars as a percentage) (Bleisch et al. 2008, p. 220, ➪ Appendix A)

3.1.1.4. Estimating absolute bar lengths

The task of estimating absolute bar lengths from displays with frames is completed with the same level of success in the 3D setting as in the 2D setting (figures 3.5 and 3.6). Applying a Student t-test shows that the mean of the differences (figure 3.5) is significantly different from 0 for the setting '2D f'. Here, the participants tend to slightly underestimate the bar lengths when responses are compared with actual values. Using a Tukey-Test for the means of the task completion times, the alternative hypothesis $H_1 : m_{2Df} \neq m_{3Df}$ is accepted with the 3D setting taking longer (note that 3 seconds of 3D scene start-up time were subtracted in the setting '3D f'). (Bleisch et al. 2008, p. 221, ➪ Appendix A)

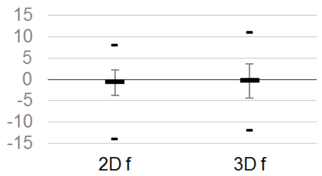


Figure 3.5.: Mean, standard deviation, and minimum/maximum values of the differences between recorded and absolute values when estimating absolute values (Bleisch et al. 2008, p. 221, ➪ Appendix A)

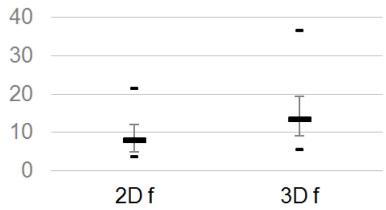


Figure 3.6.: Mean, standard deviation (calculated from log normal distribution), and minimum/maximum values of task completion time in seconds for estimating and recording absolute values (Bleisch et al. 2008, p. 221, ➪ Appendix A)

3.1.2. Testing research hypotheses

✔ H I.I: Users are able to identify the taller of two bars in a 3D desktop virtual environment as well as they can in static 2D graphics.

Informally, the evidence suggests that participants are able to identify the taller of two bars in all four settings. Thus hypothesis I.I is accepted. As shown in section 3.1.1.3, the taller bar could be identified in almost 100% of the cases in all four settings ('2D f', '2D nf', '3D f' and '3D nf'). Most errors occurred where the smaller bar was sized 86% or more of the larger bar, independent of absolute size. (Bleisch et al. 2008, p. 223f, ➪ Appendix A)

✔ H I.II: The effectiveness of estimating differences between two bars is not significantly different in the 2D and 3D settings (between '2D nf' and '3D f').

The results discussed in section 3.1.1.2 show that there are significant differences between the four settings but not between the settings '2D nf' and '3D f'. Thus, hypothesis I.II is accepted. (Bleisch et al. 2008, p. 224, ➪ Appendix A)

✘ ✔ H I.III: The efficiency of task completion is improved by the use of a reference grid (frame) (in the 2D ✘ and 3D ✔ settings).

Hypothesis I.III is rejected for the 2D settings as there were no significant differences found between the settings with and without frames in 2D. However, in the 3D setting with frames the tasks take significantly less time than in the 3D setting without frames. (Bleisch et al. 2008, p. 224, ➩ Appendix A)

✘ H I.IV: The efficiency of estimating absolute values from bar lengths (with a reference frame) is not significantly different in the 2D and 3D settings.

The results discussed in section 3.1.1.4 require that hypothesis I.IV is rejected. There were significant differences found between the 2D and 3D settings with frames. However, the results are very accurate. (Bleisch et al. 2008, p. 224, ➩ Appendix A)

3.1.3. Further analysis of the data

The data was analysed further to consider various influences of recorded characteristics that may affect performance. These include the effects of participants' general spatial ability on task performance and the task completion times, the position of the bars in the landscape, the vertical non-alignment of the bars in the 2D settings and the different combinations of bars. None of these potential influences was found to be significant (details in Bleisch et al. 2008, ➩ Appendix A). However, testing participant's general spatial ability with the Santa Barbara Sense of Direction test (Hegarty, Richardson, Montello, Lovelace & Subbiah 2002) allowed confirmation that the informed users (section 2.4) participating in this research generally have high spatial abilities as was aimed for.

3.2. Stage IIa - Evaluating single bars with more complex tasks

3.2.1. Description of the sample

The questionnaire and tasks for stage IIa were sent out to 74 persons of the potential group of informed users (section 2.4). 34 of them did the tasks and filled in the questionnaire with their answers. Information about the process and the collected data can be found in section 2.5.2.

The content of the raw data are the tasks, the 932 answers (2D: 468, 3D: 464) to them, a confidence rating of the answer and the time needed for each answer. Additionally, it is known from the methodical design (section 2.5.2) in what experimental environment, e.g. dimension (2D or 3D), setting and data set, the answers were given. Most of the answers are quite complex and comprise of several insights (section 2.6.2.1). In total there are 847 insights in 2D and 891 insights in 3D.

3.2.1.1. Time

In total the participants spent 45 hours and 11 minutes (about one working week) or Ø1h 20min per participant with the questionnaire including time to get used to the settings and environments. The total time spent fulfilling tasks and giving answers was 31 hours 21 minutes 19 seconds. Figure 3.7 shows the answer times in seconds in 2D and 3D as boxplots. The time values are not normally distributed in 2D (W = 0.781, p-value < 0.001) nor in 3D (W = 0.813, p-value < 0.001). Comparing the 2D and 3D time values (Z = -1.175, p-value = 0.240) shows that there is no significant difference between 2D and 3D.

3.2.1.2. Con dence rating

The participants confidence rating (low, medium or high) for each answer in 2D and 3D are shown in figure 3.8. Comparing the confidence ratings in 2D and 3D (χ^2 = 1.910, df = 2, p-value = 0.385) shows that there is no significant difference. The rating 'medium' shows slightly more ratings for the 2D settings. Another potential trend is that the high and low confidence ratings are slightly higher for the tasks in the 3D visualisations.

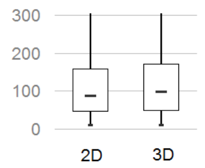


Figure 3.7.: Boxplots of 2D (left, n=468, max value=1016) and 3D (right, n=464, max value=830) times per answer in seconds

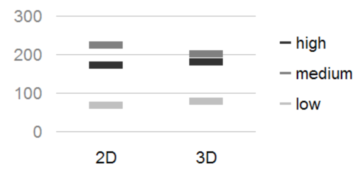


Figure 3.8.: Confidence ratings of answers in 2D and 3D (number of answers with low, medium and high confidence)

3.2.1.3. Complexity

The insights found within the task answers are rated regarding their complexity (section 2.6.2.1). Figure 3.9 shows the number of insights with low (c low), medium (c medium) and high (c high) complexity in 2D and 3D. Comparing the complexity of insights in 2D and 3D (χ^2 = 4.669, df = 2, p-value = 0.097) shows that there is no significant difference.

3.2.1.4. Plausibility

The insights found within the task answers are rated regarding their plausibility (section 2.6.2.1). Figure 3.10 shows the number of insights with low (p low), medium (p medium) and high (p high) plausibility in 2D and 3D. Comparing the plausibility of insights in 2D and 3D (χ^2 = 4.929, df = 2, p-value = 0.085) shows that there is no significant difference.

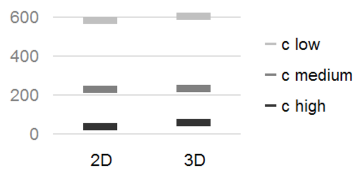


Figure 3.9.: Complexity of insights in 2D and 3D (number of insights with low, medium and high complexity)

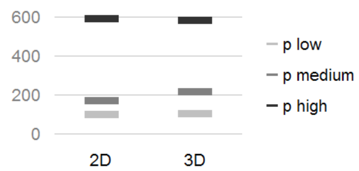


Figure 3.10.: Plausibility of insights in 2D and 3D (number of insights with low, medium and high plausibility)

3.2.1.5. Word count analysis

The task answers were analysed regarding the use of different categories of words (section 2.6.2.4 and table 2.5 for an explanation of the word categories). Figure 3.11 shows the numbers of words from categories L1-L3 and A1-A3 used in 2D and 3D. Comparing the word counts in 2D and 3D (χ^2 = 25.840, df = 5, p-value < 0.001) shows that there is a significant difference in the use of the words in 2D and 3D.

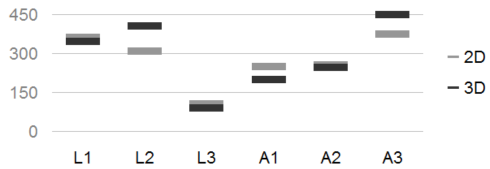


Figure 3.11.: Word counts for the different word categories L1-L3 and A1-A3 (explained in table 2.5) in 2D and 3D

3.2.2. Testing research hypotheses

3.2.2.1. Hypotheses regarding ef ciency

✘ H IIa.Ia: Tasks where the reference set is location are solved more efficiently in the 2D representation.

Comparing the task times in seconds for the two tasks with location as reference set (section 2.5.2.2) in 2D and 3D (Z = -0.755, p-value = 0.450) shows that there is no significant difference. Thus the hypothesis that tasks with reference set location are solved more efficiently in the 2D representation is rejected (figures 3.12 and 3.13).

✘ H IIa.IIa: Tasks where the reference set is altitude are solved more efficiently in the 3D virtual environment.

Comparing the task times in seconds for the two tasks with altitude as reference set (section 2.5.2.2) in 2D and 3D (Z = -0.271, p-value = 0.786) shows that there is no significant difference. Thus the hypothesis that tasks with reference set altitude are solved more efficiently in the virtual environment is rejected (figures 3.12 and 3.13).

✘ H IIa.IIIa: Tasks that include combined reference sets (location and altitude) are solved more efficiently in the 3D virtual environment.

Comparing the task times in seconds for the two tasks with altitude and location as reference set (section 2.5.2.2) in 2D and 3D (Z = -1.077, p-value = 0.281) shows that there is no significant difference. Thus the hypothesis that tasks with a combined reference set (location and altitude) are solved more efficiently in the virtual environment is rejected (figures 3.12 and 3.13).

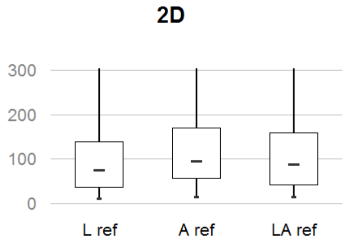


Figure 3.12.: Boxplots of task times for different reference sets (L location, A altitude and LA both) in 2D (maximum values: L=468, A=665, LA=1016)

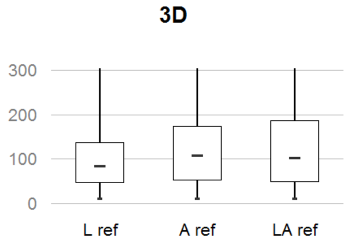


Figure 3.13.: Boxplots of task times for different reference sets (L location, A altitude and LA both) in 3D (maximum values: L=809, A=830, LA=619)

3.2.2.2. Hypotheses regarding effectiveness

✘ H IIa.Ib: Tasks where the reference set is location are solved more effectively in the 2D representation.

Complexity: Testing for a difference between 2D and 3D regarding insight complexity (low complexity, medium complexity and high complexity) for tasks with reference set location (χ^2 = 4.602, df = 2, p-value = 0.100) shows that there is no significant difference (figure 3.14).

Plausibility: Testing for a difference between 2D and 3D regarding insight plausibility (low plausibility, medium plausibility and high plausibility) for tasks with reference set location (χ^2 = 0.495, df = 2, p-value = 0.781) shows that there is no significant difference (figure 3.15).

The hypothesis that tasks with reference set location are solved more effectively in the 2D representation is rejected as there is no significant difference between 2D and 3D regarding insight complexity and plausibility.

✘ H IIa.IIb: Tasks where the reference set is altitude are solved more effectively in the 3D virtual environment.

Complexity: Testing for a difference between 2D and 3D regarding insight complexity (low, medium and high complexity) for tasks with reference set altitude (χ^2 = 3.362, df = 2, p-value = 0.186) shows that there is no significant difference (figure 3.14).

Plausibility: Testing for a difference between 2D and 3D regarding insight plausibility (low plausibility, medium plausibility and high plausibility) for tasks with reference set altitude (χ^2 = 7.629, df

= 2, p-value = 0.022) shows that there is a significant difference. The data displayed in figure 3.15 indicates that the 3D virtual environment creates more medium plausible insights compared to the 2D representation.

The hypothesis that tasks with reference set altitude are solved more effectively in the 3D virtual environment is rejected for the aspect of complexity. There is a significant difference for plausibility. But as creating more medium plausible insights may not be judged as being more efficient the hypothesis is also rejected for the aspect of plausibility.

✘ H IIa.IIIb: Tasks that include combined reference sets (location and altitude) are solved more effectively in the 3D virtual environment.

Complexity: Testing for a difference between 2D and 3D regarding insight complexity (low, medium and high complexity) for tasks with a combined reference set location and altitude (χ^2 = 3.530, df = 2, p-value = 0.171) shows that there is no significant difference (figure 3.14).

Plausibility: Testing for a difference between 2D and 3D regarding insight plausibility (low, medium and high plausibility) for tasks with a combined reference set location and altitude (χ^2 = 6.088, df = 2, p-value = 0.048) shows that there is a significant difference. The data displayed in figure 3.15 indicates that in the 3D virtual environment more low and medium plausible insights but less highly plausible insights are created.

The hypothesis that tasks with a combined reference set location and altitude are solved more effectively in the 3D virtual environment is rejected for the aspect of complexity as there is no significant difference between 2D and 3D. However, there is a significant difference between 2D and 3D for the aspect of plausibility. The hypothesis is nevertheless rejected as creating more low/medium plausible insights and less highly plausible insights in the 3D virtual environment cannot be judged as being more effective. Rather the reverse case is true. The 2D representation seems to yield more highly plausible and less low or medium plausible insights and might thus be regarded as being more effective for the analysis of the data as compared to the 3D virtual environment.

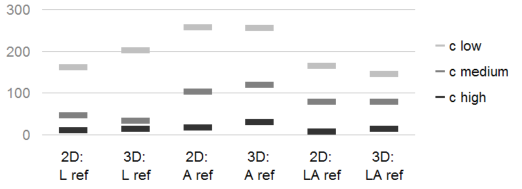


Figure 3.14.: Insight complexity per dimension (2D and 3D) and task reference set (L location, A altitude and LA both)

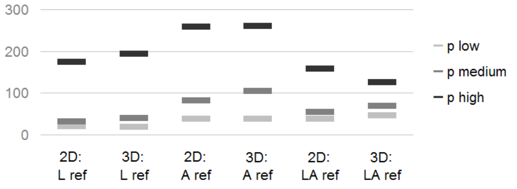


Figure 3.15.: Insight plausibility per dimension (2D and 3D) and task reference set (L location, A altitude and LA both)

3.2.3. Further analysis of the data

In addition to testing the proposed hypotheses the data of stage IIa was explored further. The following sections describe some of the findings which answer questions which arose during hypothesis testing or generally seemed interesting and potentially relevant regarding the aims of this research (section 1.4).

3.2.3.1. Complexity vs. plausibility

The evaluation of complexity and plausibility separately (sections 3.2.1.3 and 3.2.1.4) leaves open the question if there are differences between 2D and 3D in the plausibility of differently complex answers. The numbers in many cells of the contingency tables are too small to make a useful χ^2 test and aggregating cells does not make sense (section 2.6.1.3). Figure 3.16 shows informally that there are no big differences between 2D and 3D for complexity versus plausibility.

Figure 3.16.: Numbers of insights with low, medium and high complexity, separated by task reference (L location, A altitude and LA both) and dimension (2D and 3D)

3.2.3.2. Task references vs. insight references

Testing for differences in insight reference (L location, A altitude and LA both) between 2D and 3D shows that there is no significant difference either for reference location (χ^2 = 1.580, df = 2, p-value = 0.454), reference altitude (χ^2 = 1.207, df = 2, p-value = 0.547) or for insights referring to a combination of location and altitude (χ^2 = 1.285, df = 2, p-value = 0.526). The data displayed in figure 3.17 suggest that insight references are strongly dependent on task reference. Testing for differences between question reference and insight references shows that there is a significant difference for 2D (χ^2 = 290.922, df = 4, p-value $<$ 0.001) and 3D (χ^2 = 339.859, df = 4, p-value $<$ 0.001).

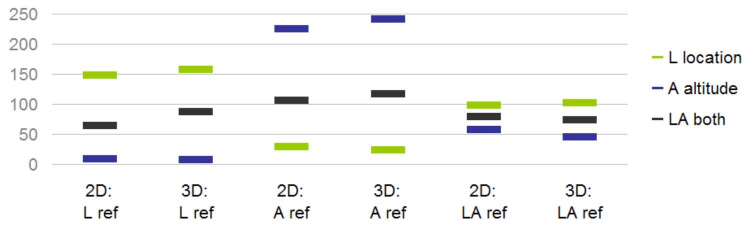


Figure 3.17.: Number of insights per reference type (L location, A altitude and LA both), separated by task reference (L, A and LA) and dimension (2D and 3D)

3.2.3.3. Con dence ratings

Task references may influence the confidence ratings for each answer given by the participants (figure 3.18). Testing for differences between confidence ratings shows a significant variation for 2D (χ^2 = 18.791, df = 4, p-value = 0.001) but not for 3D (χ^2 = 7.155, df = 4, p-value = 0.128). Additionally, no significant differences can be found between 2D and 3D either for task reference location (χ^2 = 1.0625, df = 2, p-value = 0.588), altitude (χ^2 = 2.341, df = 2, p-value = 0.310) or for location and altitude combined (χ^2 = 3.607, df = 2, p-value = 0.165).

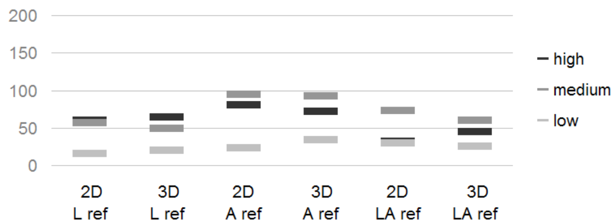


Figure 3.18.: Confidence ratings (high, medium and low) per task reference (L location, A altitude and LA both) in 2D and 3D

Assigning the different settings, dimensions and tasks in a balanced within subject design (section 2.5.2.4) shall minimise, for example, learning effects. Testing the confidence ratings for differences between the first and second displays of a specific experimental setting (figure 3.19) shows that there are significant differences in 2D (χ^2 = 8.080, df = 2, p-value = 0.018) and 3D (χ^2 = 7.153, df = 2, p-value = 0.028). The differences between 2D and 3D are not significant for the first displays shown (χ^2 = 1.763, df = 2, p-value = 0.414) but are significant for the second displays (χ^2 = 10.976, df = 2, p-value = 0.004).

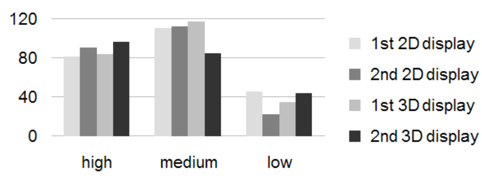


Figure 3.19.: Confidence ratings (high, medium and low) per per first and second displays in 2D and 3D

3.2.3.4. Differences between settings

This section analyses the differences between the two environmental settings W and S (section 2.5.2.3). The analysis of the differences between the eight data sets which the two settings W and S consist of is described in section 3.2.3.5.

Time: Testing for differences in time between 2D and 3D in the two environmental settings (figure 3.20) shows no significant difference either for setting W (Z = -0.269, p-value = 0.788) or for setting S (Z = -1.116, p-value = 0.264). Testing for a difference in time between the two environmental settings W and S in 2D and 3D shows no significant difference for 2D (Z = -1.275, p-value = 0.202) but a significant difference for 3D (Z = -2.010, p-value = 0.044).

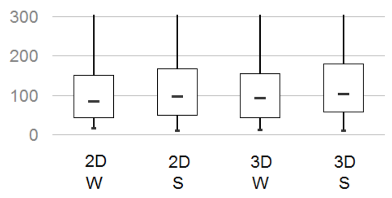


Figure 3.20.: Boxplot of times for the two environmental settings (W and S) in 2D and 3D (maximum values: 2D W=665, 2D S=1016, 3D W=830, 3D S=619)

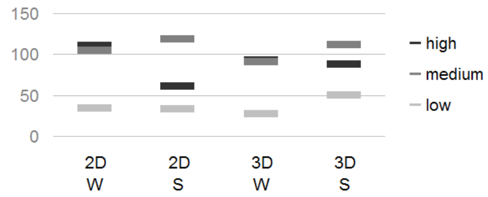


Figure 3.21.: Confidence ratings (low, medium and high) per setting (W and S) in 2D and 3D

Con dence rating: Testing for differences in the confidence ratings per answer (figure 3.21) shows no significant difference between 2D and 3D either for setting W (χ^2 = 0.090, df = 2, p-value = 0.956) or for setting S (χ^2 = 5.480, df = 2, p-value = 0.065). Between W and S in 2D (χ^2 = 12.196, df = 2, p-value = 0.002) a significant difference is found but the difference between W and S is just not significant for 3D (χ^2 = 5.919, df = 2, p-value = 0.052).

Complexity: Testing for differences in the complexity of insights (c low, c medium and c high) shows no significant difference between 2D and 3D either for setting W (χ^2 = 1.426, df = 2, p-value = 0.490) or for setting S (χ^2 = 2.508, df = 2, p-value = 0.285). Also there is no significant difference between setting W and setting S in 2D (χ^2 = 1.350, df = 2, p-value = 0.509) and 3D (χ^2 = 2.043, df = 2, p-value = 0.360) (figure 3.22).

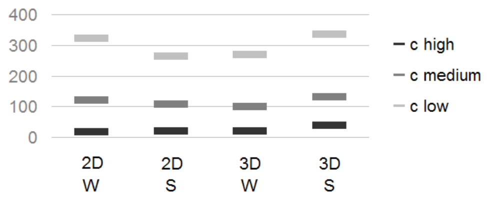


Figure 3.22.: Complexity (c low, c medium and c high) per setting (W and S) in 2D and 3D

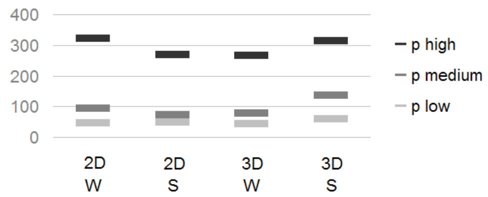


Figure 3.23.: Plausibility (p low, p medium and p high) per setting (W and S) in 2D and 3D

Plausibility: Testing for differences in the plausibility of insights (p low, p medium and p high) shows no significant difference between 2D and 3D for setting W (χ^2 = 0.308, df = 2, p-value = 0.857) but a significant difference for setting S (χ^2 = 8.266, df = 2, p-value = 0.016). There is no significant difference between setting W and setting S in 2D (χ^2 = 1.567, df = 2, p-value = 0.457) and just not for 3D (χ^2 = 5.565, df = 2, p-value = 0.062) (figure 3.23).

Insight references: Testing for differences in the references of the insights (location, L ref; altitude, A ref and location and altitude combined, LA ref) shows no significant difference between 2D and 3D either for setting W (χ^2 = 2.018, df = 2, p-value = 0.365) or for setting S (χ^2 = 0.969, df = 2, p-value = 0.616). There are significant differences between setting W and setting S in 2D (χ^2 = 18.175, df = 2, p-value $<$ 0.001) and in 3D (χ^2 = 9.268, df = 2, p-value = 0.010) (figure 3.24).

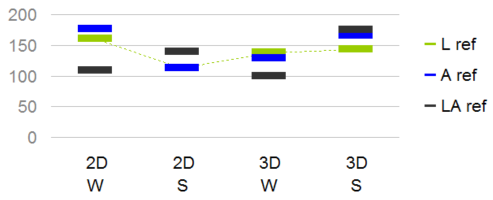


Figure 3.24.: Insight references (L location, A altitude and LA both) per setting (W and S) in 2D and 3D

3.2.3.5. Differences between data sets

This section in comparison to section 3.2.3.4 analyses the differences between the eight data sets (W1-W4 and S1-S4) which are related to the two environmental settings W and S (section 2.5.2.3). It thus details the previous section but as some numbers are small (0-113 values per category) trends may be identified but variation in numbers are not tested for significance (χ^2 testing is not suggested for small numbers, section 2.6.1.3).

Figure 3.25 shows the measured task performance time, confidence rating, complexity, plausibility and insight references for each data set (W1-W4 and S1-S4) in 2D and 3D. To compare the measures for the different data sets and to find out if potentially the 2D representation or the 3D visualisation was more useful for each of them informal ratings are applied. Each measure is assigned a $+$, $++$, $-$ or $--$ in 2D or 3D (details in table 3.1) and then the total number of $+$ and $-$ are counted for each data set in 2D and 3D. The insight references per data set (figures 3.25i and 3.25j) are not included as a rating is task and context dependent. While for some applications it may be preferred to have more answers referring to location in an other application this might not be the case. Thus the data is shown in figures 3.25i and 3.25j for visual analysis but not integrated into the rating and judgement of the visualisation type appropriateness per data set (table 3.2).

time	$+$: median is smaller
	$+$: range is smaller
confidence rating	$++$: more high confidence ratings
	$+$: more medium confidence ratings
	$--$: more low confidence ratings
complexity	$++$: more insights of high complexity
	$+$: more insights of medium complexity
plausibility	$++$: more insights of high plausibility
	$+$: more insights of medium plausibility
	$--$: more insights of low plausibility

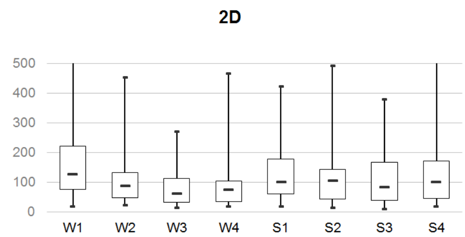
Table 3.1.: Ratings of the different measures for the comparison of the visualisations of the data sets in 2D and 3D

The result in table 3.2 shows that the 3D visualisation is most appropriate for data set S1 (mostly $+$) while the 2D representation is more appropriate for the data sets W1, W2, W3, S3, and S4 (mostly

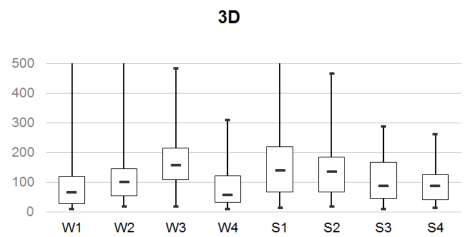
+). For the data sets W4 and S2 both visualisation types have something to offer (similar numbers of + in 2D and 3D).

	W1	W2	W3	W4	S1	S2	S3	S4
2D	x	x	x				x	x
3D					x			
2D or 3D				x		x		

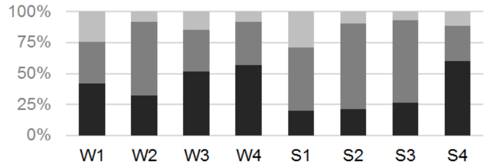
Table 3.2.: Most appropriate visualisation type per data set based on the ratings described in table 3.1



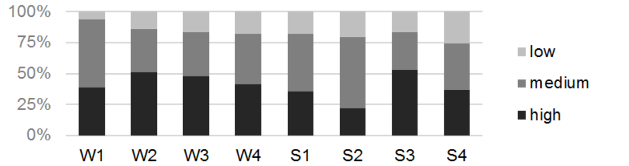
(a) Boxplots of task performance times (seconds) in 2D



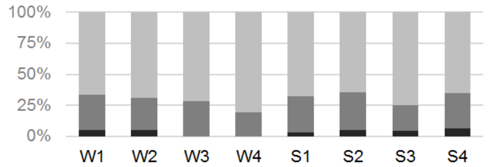
(b) Boxplots of task performance times (seconds) in 3D



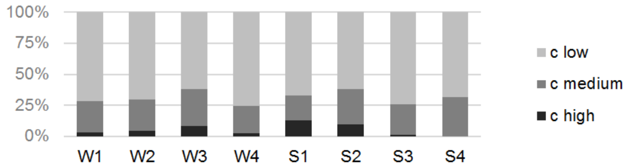
(c) Confidence ratings (low, medium and high) in 2D



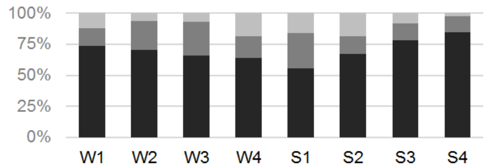
(d) Confidence ratings (low, medium and high) in 3D



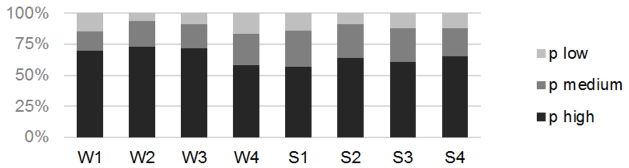
(e) Complexity (c low, c medium and c high) in 2D



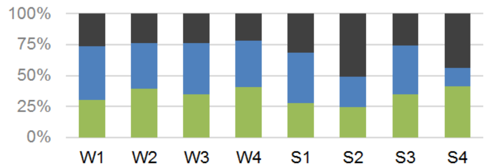
(f) Complexity (c low, c medium and c high) in 3D



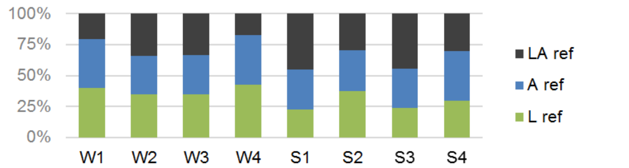
(g) Plausibility (p low, p medium and p high) in 2D



(h) Plausibility (p low, p medium and p high) in 3D



(i) Insight references (L location, A altitude and LA both) in 2D



(j) Insight references (L location, A altitude and LA both) in 3D

Figure 3.25.: The measures task performance time, confidence, complexity, plausibility and insight reference per data set (W1-W4 and S1-S4) in the settings W and S in 2D and 3D

3.2.3.6. Word count analysis

The following two sections test for differences in the quantities of words used in the categories L1-L3 and A1-A3 (section 2.6.2.4 for an explanation of the word categories).

Settings W and S: Testing for variation in the number of words used between the different setting W and S (figure 3.26) in 2D and 3D shows significant differences for setting W (χ^2 = 15.954, df = 5, p-value = 0.007) and setting S (χ^2 = 27.095, df = 5, p-value < 0.001). There is also significant variation between the settings W and S in 2D (χ^2 = 17.997, df = 5, p-value = 0.003) and in 3D (χ^2 = 46.625, df = 5, p-value =< 0.001).

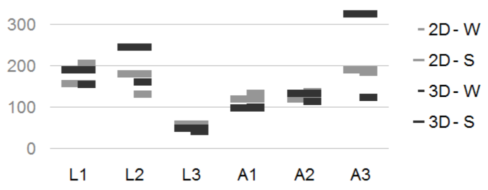


Figure 3.26.: Word counts per category L1-L3 and A1-A3 (explained in table 2.5) in the different settings W and S in 2D and 3D

Word counts vs. task references: Testing for differences in the number of words used in regard of the different tasks and thus task references (figure 3.27) shows that there is significant variation within 2D (χ^2 = 354.029, df = 10, p-value < 0.001) and 3D (χ^2 = 370.493, df = 10, p-value < 0.001). There is also significant variation between 2D and 3D for the word counts in answers to tasks with altitude as reference (χ^2 = 12.545, df = 5, p-value = 0.028) or a combined location/altitude reference (χ^2 = 13.996, df = 5, p-value = 0.016) but no significant difference between 2D and 3D for word counts in answers to tasks with location as reference (χ^2 = 8.785, df = 5, p-value = 0.118).

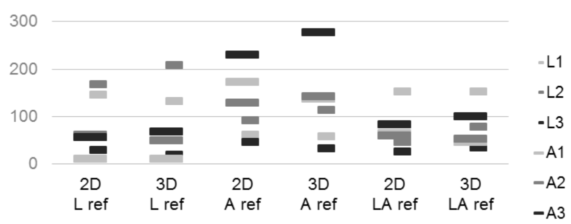


Figure 3.27.: Word counts per category (L1-L3 and A1-A3) in the answers to the tasks with different references (location, altitude and both) in 2D and 3D

3.3. Stage IIb - Evaluating bar charts with more complex tasks

3.3.1. Description of the sample

The questionnaire and tasks for stage IIb were completed by 38 participants. Information about the process and the collected data can be found in section 2.5.3.

The content of the raw data are the 522 insights (2D: 260, 3D: 262) into the data reported by the participants, a confidence rating per insight and the time needed for each insight. Additionally, it is known from the methodical design (section 2.5.3.4) in what experimental environment, e.g. dimension (2D or 3D) and data sets, the answers were given.

3.3.1.1. Time

In total the participants spent about 38 hours or Ø1h per participant with the questionnaire including time for instructions and getting used to the settings and environments. The total time spent finding answers was 15 hours 42 minutes 55 seconds. Figure 3.28 shows the answer times in seconds in 2D and 3D as boxplots. The time values are not normally distributed in 2D (W = 0.742, p-value $<$ 0.001) nor in 3D (W = 0.820, p-value $<$ 0.001). Comparing the 2D and 3D time values (Z = -0.236, p-value = 0.814) shows that there is no significant difference between 2D and 3D.

3.3.1.2. Con dence rating

The participants confidence rating (low, medium and high) for each insight in 2D and 3D are shown in figure 3.29. Comparing the confidence ratings in 2D and 3D (χ^2 = 1.332, df = 2, p-value = 0.514) shows that there is no significant difference.

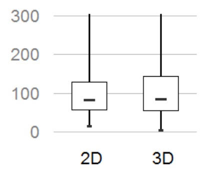


Figure 3.28.: Boxplots of 2D (left, n=260, max value=764) and 3D (right, n=262, max value=549) times per insight in seconds

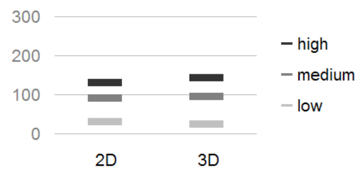


Figure 3.29.: Confidence ratings of answers in 2D and 3D (number of answers with low, medium and high confidence)

3.3.1.3. Complexity

The insights reported are rated regarding their complexity (section 2.6.2.1). Figure 3.30 shows the number of insights with low (c low), medium (c medium) and high (c high) complexity in 2D and 3D. Comparing the complexity of insights in 2D and 3D (χ^2 = 1.167, df = 2, p-value = 0.558) shows that there is no significant difference.

3.3.1.4. Plausibility

The insights reported are rated regarding their plausibility (section 2.6.2.1). Figure 3.31 shows the number of insights with low (p low), medium (p medium) and high (p high) plausibility in 2D and 3D. Comparing the plausibility of insights in 2D and 3D (χ^2 = 1.037, df = 2, p-value = 0.595) shows that there is no significant difference.

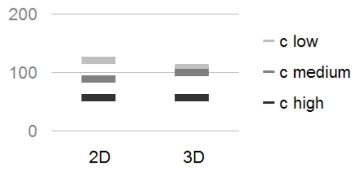


Figure 3.30.: Complexity of insights in 2D and 3D (number of insights with low, medium and high complexity)

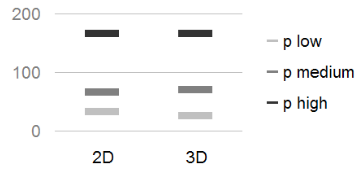


Figure 3.31.: Plausibility of insights in 2D and 3D (number of insights with low, medium and high plausibility)

3.3.1.5. Word count analysis

The insights were analysed regarding the use of different categories of words (section 2.6.2.4 and table 2.5 for an explanation of the word categories). Figure 3.32 shows the numbers of words from categories L1-L3 and A1-A3 used in 2D and 3D. Comparing the word counts in 2D and 3D (χ^2 = 14.781, df = 5, p-value = 0.011) shows that there is a significant difference in the use of the words in 2D and 3D.

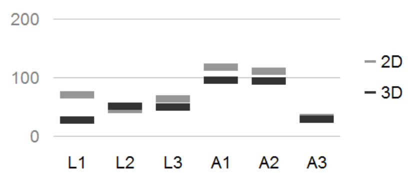


Figure 3.32.: Word counts for the different word categories L1-L3 and A1-A3 (explained in table 2.5) in 2D and 3D

3.3.2. Testing research hypotheses

3.3.2.1. Hypotheses regarding ef ciency

- ✘ H IIb.Ia: Insights into the relationship between data and location/land-cover/etc. are more often and more efficiently reported in the 2D representation.
- ✘ H IIb.Ib: Insights into the relationship between data and altitude/landform are more often and more efficiently reported in the 3D virtual environment.

Quantity: Testing for a difference between 2D and 3D regarding the number of insights with different references such as location (L ref), altitude (A ref) or location and altitude (LA ref) (χ^2 = 1.653, df = 2, p-value = 0.438) shows that there is no significant difference (figure 3.33).

Time: Testing for differences between 2D and 3D regarding the times in seconds per insight shows neither for reference location (Z = -0.205, p-value = 0.838), nor for reference altitude (Z = 0.154, p-value = 0.877) nor for reference location and altitude (Z = -0.304, p-value = 0.761) a significant difference (figures 3.34 and 3.35).

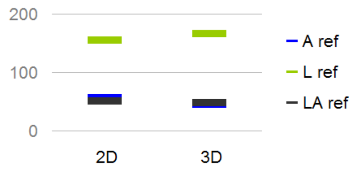


Figure 3.33.: Number of insights per reference (L location, A altitude, LA both) and dimension (2D and 3D)

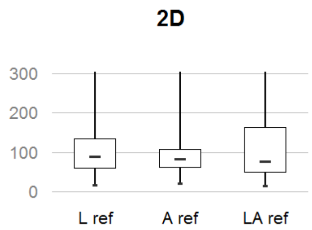


Figure 3.34.: Boxplots of insight times for different reference sets (L location, A altitude, LA both) in 2D (maximum values: L=537, A=411, LA=764)

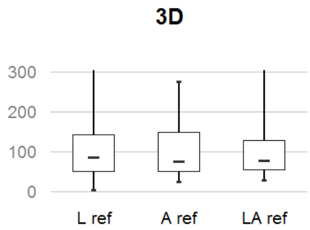


Figure 3.35.: Boxplots of insight times for different reference sets (L location, A altitude, LA both) in 3D (maximum values: L=549, A=274, LA=408)

The two hypotheses IIb.Ia and IIb.Ib stating that insights into the relationship between data and reference location are more often and more efficiently reported in the 2D representation and the insights into the relationship between data and reference altitude are more often and more efficiently reported in the 3D virtual environment are both rejected. Neither the quantity nor the insight times show a significant difference between 2D and 3D.

3.3.2.2. Hypotheses regarding effectiveness

✔ H IIb.II: The complexity and plausibility of insights reported do not vary significantly between the 2D representation and the 3D virtual environment.

Complexity: Testing for a difference between 2D and 3D regarding the complexity of insights with different references (L location, A altitude, LA both) shows no significant differences for location L (χ^2 = 0.311, df = 2, p-value = 0.856), altitude A (χ^2 = 0.847, df = 2, p-value = 0.655) or for location and altitude combined LA (χ^2 = 2.166, df = 2, p-value = 0.339) (figure 3.36).

Plausibility: Testing for a difference between 2D and 3D regarding the plausibility of insights with different references (L location, A altitude, LA both) shows no significant differences for location L (χ^2 = 0.147, df = 2, p-value = 0.929), altitude A (χ^2 = 0.952, df = 2, p-value = 0.621) or for location and altitude combined LA (χ^2 = 2.772, df = 2, p-value = 0.250) (figure 3.37).

The hypothesis that the complexity and plausibility of insights reported do not vary significantly between 2D and 3D can thus be accepted. Neither the complexity nor the plausibility show a significant difference between 2D and 3D for any of the three reference types (location L ref, altitude A ref and combined LA ref).

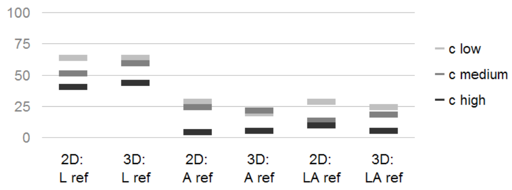


Figure 3.36.: Complexity of insights (low, medium and high) per reference (L location, A altitude, LA both) and dimension (2D and 3D)

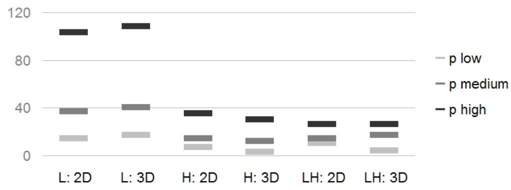


Figure 3.37.: Plausibility of insights (low, medium and high) per reference (L location, A altitude, LA both) and dimension (2D and 3D)

3.3.3. Further analysis of the data

In addition to testing the proposed hypotheses the data of stage IIb was further explored. The following sections describe some of the findings which answer questions which arose during hypothesis testing or generally seemed interesting and potentially relevant regarding the aims of this research (section 1.4).

3.3.3.1. Complexity vs. plausibility

The evaluation of complexity and plausibility separately (section 3.3.2.2) leaves open the question of whether there are differences between 2D and 3D in the plausibility of differently complex answers (figure 3.38). Testing shows that there are no significant differences between 2D and 3D either for highly plausible insights (χ^2 = 2.036, df = 2, p-value = 0.361), medium plausible insights (χ^2 = 0.144, df = 2, p-value = 0.931) or low plausible insights (χ^2 = 2.383, df = 2, p-value = 0.304).

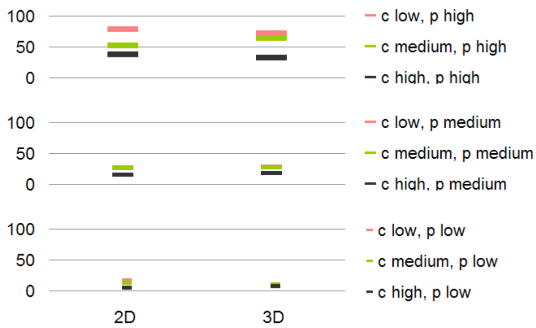


Figure 3.38.: Complexity for low, medium and highly plausible insights in 2D and 3D

3.3.3.2. Differences between data sets and settings

Stage IIb consists of two different data sets in two different environmental settings W and S (section 2.5.3.3 for details).

Time: Testing for differences in time between 2D and 3D in the two environmental settings shows no significant difference either for setting W (Z = 0.560, p-value = 0.575) or for setting S (Z = -0.849, p-value = 0.396). Testing for a difference in time between the two environmental settings W and S (figure 3.39) shows no significant difference either for 2D (Z = -0.775, p-value = 0.439) or for 3D (Z = 0.502, p-value = 0.615).

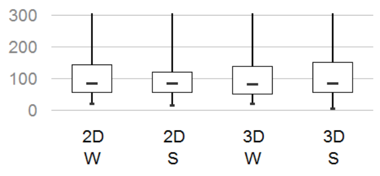


Figure 3.39.: Boxplot of times for the two environmental settings (W and S) in 2D and 3D (maximum values: 2D W=537, 2D S=764, 3D W=472, 3D S=549)

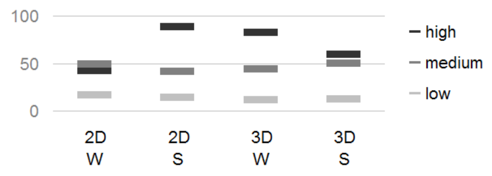


Figure 3.40.: Confidence ratings (low, medium and high) per setting (W and S) in 2D and 3D

Con dence rating: Testing for differences in the confidence ratings per answer (figure 3.40) shows a significant difference between 2D and 3D for setting W (χ^2 = 10.569, df = 2, p-value = 0.005) but not for setting S (χ^2 = 4.966, df = 2, p-value = 0.083). Between W and S in 2D (χ^2 = 12.190, df = 2, p-value = 0.002) a significant difference is found but the difference between W and S is not significant for 3D (χ^2 = 3.198, df = 2, p-value = 0.202).

Complexity: Testing for differences in the complexity of insights (c low, c medium and c high) shows no significant difference between 2D and 3D neither for setting W (χ^2 = 1.467, df = 2, p-value = 0.480) nor for setting S (χ^2 = 1.088, df = 2, p-value = 0.581). Also there is no significant difference between setting W and setting S in 2D (χ^2 = 1.370, df = 2, p-value = 0.504) and 3D (χ^2 = 2.343, df = 2, p-value = 0.310) (figure 3.41).

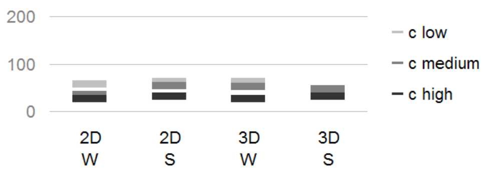


Figure 3.41.: Complexity (c low, c medium and c high) per setting (W and S) in 2D and 3D

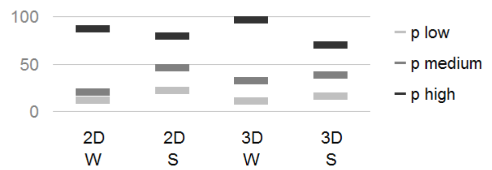


Figure 3.42.: Plausibility (p low, p medium and p high) per setting (W and S) in 2D and 3D

Plausibility: Testing for differences in the plausibility of insights (p low, p medium and p high) shows no significant difference between 2D and 3D neither for setting W (χ^2 = 1.648, df = 2, p-value = 0.439) nor for setting S (χ^2 = 0.287, df = 2, p-value = 0.866). There is a significant difference between setting W and setting S in 2D (χ^2 = 10.161, df = 2, p-value = 0.006) but not for 3D (χ^2 = 4.967, df = 2, p-value = 0.083) (figure 3.42).

Insight references: Testing for differences in the references of the insights (L location, A altitude and LA location and altitude combined) shows no significant difference between 2D and 3D neither for setting W (χ^2 = 4.616, df = 2, p-value = 0.099) nor for setting S (χ^2 = 0.327, df = 2, p-value = 0.849). There is no significant difference between setting W and setting S in 2D (χ^2 = 2.629, df = 2, p-value = 0.269) but the difference is significant in 3D (χ^2 = 7.613, df = 2, p-value = 0.022) (figure 3.43).

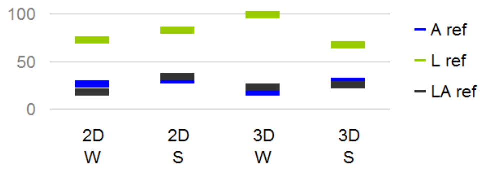


Figure 3.43.: Insight references (L location, A altitude and LA both) per setting (W and S) in 2D and 3D

3.3.3.3. Word count analysis

Settings W and S: Testing for variation in the number of words used between the different setting W and S (figure 3.44) in 2D and 3D shows just no significant difference for setting W (χ^2 = 10.703, df = 5, p-value = 0.058) but a significant variation for setting S (χ^2 = 31.504, df = 5, p-value $<$ 0.001). There is also significant variation between the settings W and S in 2D (χ^2 = 13.287, df = 5, p-value = 0.021) and in 3D (χ^2 = 20.280, df = 5, p-value = 0.001).

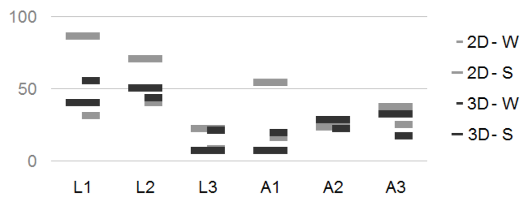


Figure 3.44.: Word counts per category L1-L3 and A1-A3 (explained in table 2.5) in the different settings W and S in 2D and 3D

3.4. Comparing stages IIa and IIb

3.4.0.1. Time

The total time spent is split according to the number of answers and insights recorded in stage IIa. This allows a rough comparison of insight finding times between stages IIa and IIb. The mean time spent for finding and reporting an insight is longer in stage IIb than in IIa (1.6 times the time for 2D and 1.7 times the time for 3D, figure 3.45).

3.4.0.2. Con dence

Testing for differences in the participants' confidence into their reported answers/insights (figure 3.46) shows that there is no significant difference between the stages IIa and IIb neither for 2D (χ^2 = 2.258, df = 2, p-value = 0.323) nor 3D (χ^2 = 2.912, df = 2, p-value = 0.233).

3.4.0.3. Complexity

Testing for differences in the complexity of insights between the two stages IIa and IIb (figure 3.47) shows that there is a significant difference for 2D (χ^2 = 8.528, df = 2, p-value = 0.014) and for 3D (χ^2 = 8.510, df = 2, p-value = 0.014).

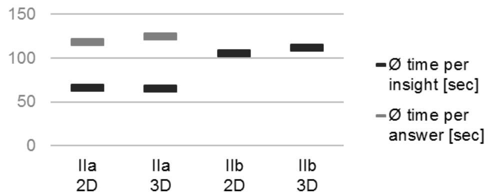


Figure 3.45.: Mean time in seconds per answer and insight for IIa and IIb in 2D and 3D

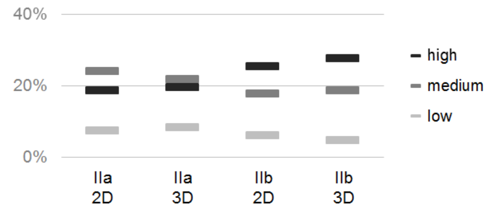


Figure 3.46.: Confidence ratings for IIa and IIb in 2D and 3D

3.4.0.4. Plausibility

Testing for differences in the plausibility of insights between the two stages IIa and IIb shows that there is no significant difference neither for 2D (χ^2 = 0.503, df = 2, p-value = 0.778) nor for 3D (χ^2 = 0.202, df = 2, p-value = 0.904) (figure 3.48).

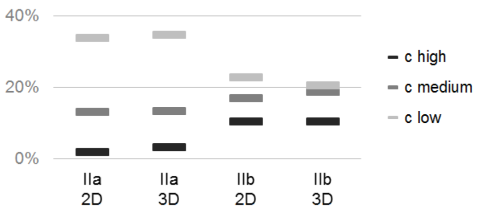


Figure 3.47.: Complexity of insights for IIa and IIb in 2D and 3D

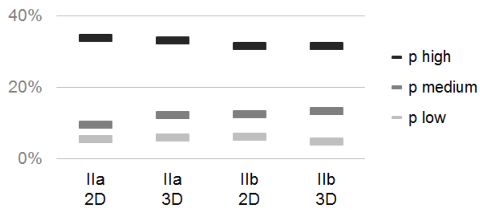


Figure 3.48.: Plausibility of insights for IIa and IIb in 2D and 3D

3.4.0.5. References of insights

Testing for differences in the quantity of insights referring either to location (L), altitude (A) or both (LA) between the two stages IIa and IIb (figure 3.49) shows that there is a significant difference for 2D (χ^2 = 6.443, df = 2, p-value = 0.040) and 3D (χ^2 = 9.847, df = 2, p-value = 0.007).

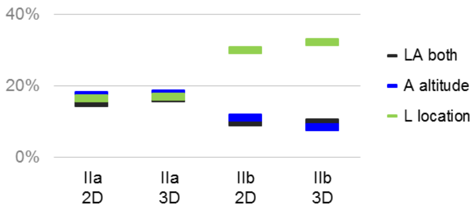


Figure 3.49.: References (L, A and LA) of insights for IIa and IIb in 2D and 3D

3.4.0.6. Word count analysis

Testing for differences in the quantity of words in the categories L1-L3 and A1-A3 (section 2.6.2.4) between the two stages IIa and IIb (figure 3.50) shows that there is just no significant difference for

2D (χ^2 = 10.683, df = 5, p-value = 0.058) but a significant difference for 3D (χ^2 = 15.175, df = 5, p-value = 0.010).

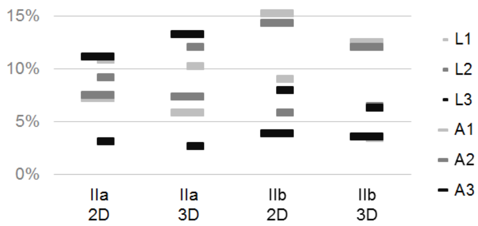


Figure 3.50.: Word counts in the categories L1-L3 and A1-A3 (explained in table 2.5) for IIa and IIb in 2D and 3D

3.5. Stage III - Evaluating bar charts with more complex tasks and application context

3.5.1. Data experts characteristics

The three case study data experts (data experts I, II and III) answered the 20 different bar length comparison tasks of stage I (section 2.5.1) in the virtual environment with frame data sets (section 2.5.4). Data experts I and II informally reported that the found it somewhat difficult to compare the two bars in the virtual environment and that they are not too confident in their answers. Figure 3.51 shows that the medians of bar length comparison errors for all three case study experts are just within the middle 50% (light grey band) of the data from stage I. The second task of identifying the taller of two bars is done correctly by the data experts I and II but less exactly, but not incorrect (as defined in Bleisch et al. (2008)), by data expert III (figure 3.52). For task times, figure 3.53 shows that the medians of the task times for data experts II and III are within the middle 50% (light grey band) of the times from stage I. Only data expert I takes more time but with less spread. But the task time values of data expert I are still within the range of the stage I task times values. Based on these results the case study data experts are considered to belong to the group of informed users who participated in the previous stages. Some variation in this group is expected as rather than aiming for a homogenous group of participants a group of potential or informed users is the objective (section 2.4).

3.5.2. Summary of the case data

The thick matrix displays (section 2.6.2.5) in tables 3.3, 3.4, 3.5 and 3.6 show the condensed and summarised data set resulting from the three case studies with data experts. The case study method and the case study data sets and visualisations from three different applications areas are detailed in section 2.5.4 (table 2.3 for the data sets and case characteristics).

The data displayed in table 3.3 sets the case study scenes by giving background and context information for each case. It describes why the data was originally collected, what typical research questions are and how the data has been analysed so far. Table 3.4 contains the case study results

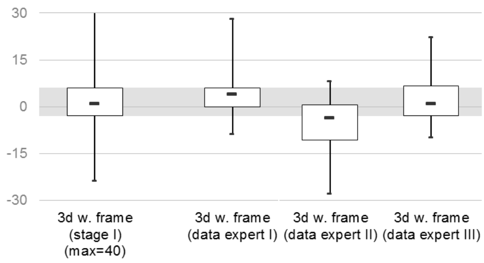


Figure 3.51.: Boxplots of the bar length comparison values in the setting 3D with frames of the stage I participants and the three case study data experts (data experts I, II and III)

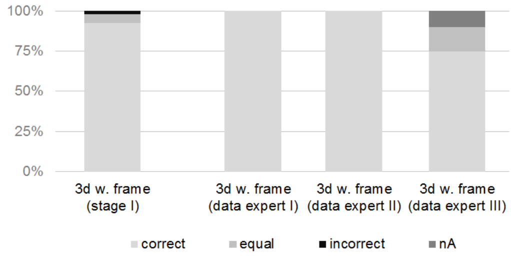


Figure 3.52.: Comparison of correct identification of the taller bar between all of the stage I participants and the three case study data experts (data experts I, II and III)

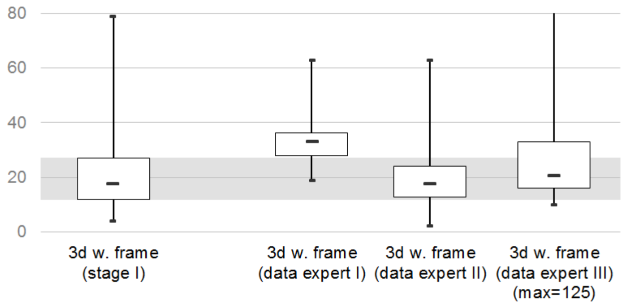


Figure 3.53.: Boxplots of task times in seconds of all the stage I participants and the three case study data experts (data experts I, II and III)

regarding the data experts use, analysis and understanding of the data. They explain their data sets, point to interesting parts of the data and explain typical data analysis while using the provided 3D visualisations of their data set(s). Table 3.5 summarises further input of the data experts regarding their familiarity with the data, comments regarding the visualisation and different uses of visualisations and a summarising appraisement of the data experts interest in the tested visualisations. Finally, table 3.6 gives a graphical overview of the suitability ratings the data experts have given the visualisations. The descriptions in these tables form the basis for the subsequent analysis of the data in regard to the propositions and exploratory questions in section 3.5.3.

3.5.3. Data analysis regarding the propositions and exploratory questions

✘ P III.I: Displays of quantitative data within virtual environments are more appropriate for the visual analysis of the data if the data and landscape (2D/3D) relationship is more important in an application.

The data experts judge the data-landscape relationship in 2D (location/distribution) to be important in all three cases. The relationship of data and landscape in 3D (altitude/landform) is judged highly important in case ’Brienz’ and ’SNP’ and of medium-high importance in case ’Literature Atlas’. Looking at the typical research questions several of them relate the data to the three-dimensional landscape, for example in case ’Brienz’ the main research question is inherently three-dimensional. Additionally, the geologists hypothesis about slope movement is directly based on the measured data and the topographic characteristics of the area. In the case ’Brienz’ also local data and landscape rela-

cases / aspects	case 'Brienz'	case 'Literature Atlas'	case 'Deer SNP'
Why was the data collected?	monitoring of a landslide area	demonstration of the potential of literature geographic methods in a concise case study	continuous research programme in the Swiss National Park (SNP) - ungulates in an alpine habitat
current exploration and analysis of the data set	calculate difference vectors in position and height between coordinate values determined in two subsequent years	queries in database, automatic creation of maps and visualisations, mainly statistical surfaces (for single texts or several texts), for visual analysis	Geographic Information System (GIS), dependent on the research questions
current visualisations and displays employed	listings, maps with overlayed point movement vectors and some visualisation 'experiments' (G2007 2010)	manually created maps (InDesign) for the central Switzerland data set (Piatti 2008) and statistical surfaces (JavaScript) for the northern Germany data set (examples in Piatti, Bär, Reuschel & Hurni 2008)	coloured points and areas on maps and ortho imagery (examples in Meyer & Filli (2006) and Campell & Filli (2006))
reports of the data analysis	internal technical report and a report for the client for each measured epoch and a website (G2007 2010) for general information	central Switzerland data set (Piatti 2008); northern Germany data set: publication of preliminary results in Piatti et al. (2008)	in Filli & Suter (2006)
Is the data related to the landscape? Is this important?	The relation of the data to topography and land-cover (e.g. road, grassland or forest) is especially important for the plausibility evaluation of the difference vectors.	Literature references the existing space (also for creating imaginative spaces), central Switzerland: literature refers directly to mountains, valleys, lakes or glaciers; references can be interpreted based on their thematic contents.	relation between data and landscape is important for understanding the habitat selection of red deer during different times of the year (e.g. Spring: May, Summer: Jun-Oct, Autumn: Nov-Dec and Winter: Jan-Apr)
typical research questions or tasks	Which point (where) moves in what direction (three-dimensional) and how much? show the movement behaviour of the landslide area in space and time; numerical testing for statistical significance of the difference vectors (significant = slope movement, insignificant = measurement noise); graphical plausibility evaluation of the difference vectors in comparison to slope inclination and land-cover; evaluation of multi epochs, shows movement trends over the years (slower, constant, faster)	Where is the setting of literature? Where do literature scenes cluster and why? Where are the blank spaces - the 'unwritten' regions? How does the fictionalisation of space develop over time? Where and when do which landscapes and cities emerge on the literary map of Europe, and when are they submerged again in meaninglessness, or when have they exhausted their literary potential? In what different ways do literary texts use the real space or - in fiction - transform the real space? How internationally is a space occupied? Or is the space inscribed almost exclusively by native authors? (Piatti 2007)	generally, migration of red deer; examples: "What is the size of the ranges used by hinds in summer and in winter? Where and how far do hinds migrate from their Il Fuorn and Val Trupchun ranges?" (Meyer & Filli 2006, p. 82) "What are the activity budgets of red deer that remain on open ground during the day instead of retreating into forest cover? Is there a difference in behavioural patterns and grazing intensity among red deer in various areas of the SNP?" (Schütte-Krug & Filli 2006, p. 106) "Do habitat factors, such as aspect, slope inclination, elevation above sea level and vegetation cover, affect winter home range selection?" (Campell & Filli 2006, p. 119)
difficulties (so far)	preparation of suitable visualisations for data and result communication to lay users (client), suitable integration of all data values into the visualisations	(in)homogeneous data collection (interpretation of literature by different researchers); visualisations: data density/symbol overlaps; scales (regional to global), texts are set on one, some or all scales	showing / analysing deer movement paths
important case study data which is not visualised	bearing of the difference vectors point numbers (e.g. as labels) 5 year sections are ok, largest shift between 1999 and 2000, was the last epoch with large shifts; this is not visible	areal or path data uncertainty/fuzziness of data and development over time read exact values from bars, e.g. by further subdividing the reference frame (scale)	deer movement (path data)

Table 3.3.: Context information and descriptive case data (table data are condensed case study data experts views or from referenced literature); table 2.3 for a general description of the cases and data sets

case 'Brienz'	case 'Literature Atlas'	case 'Deer SNP'
steep slope (Brunni), alp (Aegerdi) above flatter, Aegerdi: small shifts, in the slope/shifting area where it is steep: faster/larger shifts differences in heights: where the positional shift is large there the height shift is small and vice versa ⇒ this shows the shifting attitude of the whole slope, slope area Brunni sags (positional displacement rather than height difference, especially in the lower part) ⇒ Aegerdi slides after it; this supports the step theory of the underground, the shape of the surface may be similar to the sliding ground deep inside (thesis of geologists: no smooth sliding surface in the underground but rather steps ⇒ faster movement along vertical faces, slower movement horizontally, banking up) further down is the ”Sperre” an artificial building for keeping the slope back (gravity dam), the measured points there show more positional shifts than height shifts (especially at the top of the dam), the dam tilts slowly over because of the pressure of the slope ⇒ less differences in height; this contrary to the movements of the natural surface of the slope 2005 large storm, heavy rain fall ⇒ saturated soil; but atypical attitude of the shifting zone ⇒ no larger movement values of the last 5 years much smaller than during the beginning of the measurements ⇒ this is visible from the colours within the bars; first 10 years large shifts, later clearly decreasing shift values; reason is not known visualisation of the whole period: movement in the last 20 years (1989-2009), no subdivision there are only points and movement information where there is no forest or other difficult terrain (e.g. no point in the scree area ”In Brichen”), in such areas different measurement methods would be needed (e.g. airborne laserscanning) points/charts at the boundary of the area don't move much ⇒ they are fixed points, on these the whole adjustment calculus of the measurements are based positive height differences are not expected, they have different reasons (e.g. trigonometric height measurements ⇒ adjustment of the old height value) comparison of the data sets with the 3D landscape and the land-cover (ortho imagery) is useful	Gotthard area *[swiss data set]*: e.g. Tell' blocks some areas, but there is also free fiction possible, the landscape is multifaceted enough, has enough variety to allow this; it's also one of the rare landscape where subterranean action is possible – this is also possible in Prague (in basements, cellars and tunnels) the type of display is new but it is readily visible where there are concentrations of values, in valleys, on mountains, along the lake shore,. . . ; the information about valleys, mountains etc. could also be read from a paper map (2D), however it is more complex especially if the region is not well known change to the northern Germany data sets, p = projected (e.g. childhood area, dreaming of areas,. . .), s = place/scene of action; there are almost only imported spaces in this area last visualisation northern Germany: very interesting (zooming in to Husum); can make about the same statements as in 2D, Husum has the highest density of data, slightly more p than s, we see from this which region is more suitable for s or p, however the data outside Husum is rather exemplary and not many statements are possible yet its a good visualisation, good overview important questions in the northern Germany landscape, are there differences between center and periphery, land/shore/inland; there is no obvious need for 3D (all flat), need for 3D is more obvious in the Gotthard area, there it is important to see if the data is in valleys, on mountains or even subterranean; subterranean are especially the ”Höllloch” *[famous cave]*, the Gotthard tunnel and some smaller caves in the area northern Germany is extremely flat but already in Prague, 3D would be very interesting; there are several hills which play important roles in the literature; 3D in the urban area would be especially interesting as it has not been done before; generally a 2D city map/ortho imagery or similar is used; but especially Prague has a very distinctive topography interpretation of heights, height differences is quite easy in this visualisation	deer 604 (2004) all recordings are above the forest line in autumn and in the woods during winter time from the different distributions in the different years its possible to make statements such as which winter was harder for the red deer so far for our analysis location/movement was most often more important than altitude/landform; we do look sometimes for vertical movements of the deer (in GIS), but this is most often not obvious while only looking briefly at the visualisation interesting that the deer kept only to the left side of the valley (west of Il Fuorn); *seemed surprising* time of the day visualisations (deer 635): we did time of the day analysis sometimes, not very often, it depends on the research questions it is possible to see patterns, evening on the grassy patch, evenings and nights high up,. . . ; interesting compares visualisation to own graphics (created in ArcGIS) ⇒ it is definitively not the same, not the same impression of landscape and altitude differences; this is much better seen in Google Earth the heights/landform are normally not looked at this way, very good; gives a very good impression of top/bottom, steep, etc.; not only, near the village or in the woods, but also is it steep wood,. . . I really like this type of visualisations to look at the data and the landscape; I know the area, I am able to read maps, but to look at a visualisation like this is something else, it is useful

Table 3.4.: The expert's analysis and explanations of their data using the visualisations (table data are condensed case study data experts views, researcher's *interpretations* or *[explanations]* are written in italic type)

cases / aspects	case 'Brienz'	case 'Literature Atlas'	case 'Deer SNP'
familiarity with data	data shown in the visualisation gives the message already known; this knowledge can be read from the 3D visualisation (with bar charts); visualisations with 5 year periods contain more information than summarised visualisation; data set is very well known, maybe I am too much the data expert	data has not been seen like this before but data/area are very well known (collected and analysed by me); results are known ⇒ difficult so find/see something new Gotthard area analysed; data was digitised later, was used for first visualisations in the literature atlas project; northern Germany area: data is now being collected and simultaneously visualised for analysis/evaluation	research questions come from bigger research projects, some within the SNP only, other's in collaboration with other national parks in other countries (also comparison of similar data between different national parks); data of deers 604 and 635 is at present no longer analysed, but the data is stored and sometimes we come back to it
visualisations for analysis vs. communication purposes	effort needed to interpret point based visualisation to areal impression (in difference to e.g. interpolated surfaces) interesting to use these visualisations as communication tools, e.g. show them to the 'Gemeinderat' and see how they think about it; data experts know all the background, know how the data was collected, when something special happened etc.	in this type of visualisation (bars, point-based) ⇒ more difficult to see data patterns, see how the landscape is filled/used by literature; the literature atlas 3D display is easier to interpret; bars need interpretation of the data carpet/3D surface by the analyst; also more difficult to see regions not used by literature; literature atlas visualisations are for data experts, communication of facts by visualisations is not priority in northern Germany some new insights were found through the literature atlas visualisations; the analysis/evaluation of the data stops not with the data collection, visualisations are used	detailed analysis is done in GIS (overlays of data and other information); overlay ortho imagery with data points ⇒ see on what vegetation type the deer was; for gaining overview, seeing the story or illustration you would use such a visualisation really nice to see the 3D landscape, easier to think about it (steepness, up/down,...), especially also for presentations (show your data/findings to others who know the area less); but also useful for me (look for trends, get ideas for detailed analysis); maybe interesting to show it to data experts from other national parks (do know the type of data but not the specific landscape)
visualisation comments	black frames are not intuitive, unclear what they mean and how they are to interpret colours of the bars are nice legends of the visualisations are wrong: should say 1990 instead of 1989, first measurement was done in 1989 but the first difference value is from year 1990 (1990 values minus 1989 values)	short bars not well visible (irritating, less interpretable, diagrams overlap); easier to look for tall bars; violet hues of bars are nice but not easily interpreted (on ortho imagery background), flashy colours ⇒ better readability; white reference frames may help (too disturbing?); 6 variables in one chart are too much (for literature researcher), show maximally 3 variables at once; better stacked bars 'Tell areas' wrongly labelled (legend); legend colours seem darker than diagrams; include interface to select different themes/visualisations; spatial aggregation ev. problematic (already aggregated data in Luzern); different aggregations ⇒ diff. interpretations?	switching between all deer 604 visualisations is useful; there are symbol overlaps, but it is nevertheless useful, zoom in for details time of the day bar charts are too complex, rather have single bars showing evening, night, etc. and then playing around (switching on/off) with the different single bar visualisations aggregation to 0.001° or 0.002° WGS84 makes a difference, 0.001° more details visible, some bar graphics are splitted (single bars); 0.002° ok instead of time of the day charts (3/6 hours), preferable times of the day as night, twilight (2h), day, twilight (2h) (different hours of the day depending on the season)
summary of data experts interest (appraisement)	*medium*, data analysis traditionally, mainly numerically supported by maps and data graphics; tested visualisations may have some value for communicating results supplied visualisations were not looked at after the interview	*high*, data is already analysed using visualisations, tested visualisations are thought promising (to be included in literature atlas visualisations?) no new insights, but visualisations look promising, conceivable that insights into a new/uninterpreted data set may be gained by using such visualisations supplied visualisations were looked at after the interview and discussed with literature atlas project partners	*high*, possibilities for visual analysis with GIS systems in regard to the landscape are perceived as being complex; it is stated that this type of visualisations is very interesting for the visual analysis of deer data sets supplied visualisations were looked at after the interview and also shown to a research colleague (who thought the visualisations to be very interesting too)

Table 3.5.: Case data regarding data and visualisation use (table data are condensed case study data experts views or from referenced literature, researcher's *interpretations* or *[explanations]* are written in italic type)

question / cases	How suitable is this type of visualisation for the exploration of your data?	Do you like the visualisation?	Would you like to use this type of visualisation for other similar data sets?
case 'Brienz'			
case 'Literature Atlas'			
case 'Deer SNP'			
	very suitable … not suitable	very much … not at all	very much … not at all

Table 3.6.: Suitability ratings of the visualisations by the data experts (rated on a seven-point scale)

tionship characteristics such as "towards the bottom of the slope" may play an important role in the data analysis. In case 'Deer SNP' the research questions ask for the evaluation of 2D distribution of deer with some three-dimensional aspects and specific research questions relating data to landform are asked. Based on these research questions the data of the case 'Deer SNP' has so far only occasionally been analysed in regard to altitude and landform. More often the relation of data to location and 2D movement was analysed. However, the 3D visualisations provide easy comparison of data with altitude and landform and may thus more often be used for this type of analysis in the future. In the case 'Literature Atlas' the typical research questions show a more general interest in the landscape also potentially at a larger scale. Nevertheless, analysing data "along a mountain ridge" or "in a valley" is important especially also for the central Switzerland data set. Additionally, in the case 'Literature Atlas' the idea of visually combining data and landscape is even thought further as it could be extended to subterranean objects such as caves or tunnels and to urban areas with characteristic topography such as Prague or also to combining data with 3D city models. Looking at the data analysis in table 3.4 the data expert of case 'Brienz' reports mainly three-dimensional aspects. The data experts of the cases 'Literature Atlas' and 'Deer SNP' report 2D and 3D analysis of the data sets.

Concluding according to the proposition, displays of quantitative data within virtual environments would be most appropriate for the visual analysis of the data in the case 'Brienz' as the relationship between data and landscape is most prominent and research question and data analysis are inherently three-dimensional. They would be somewhat appropriate for the case 'Deer SNP' and least appropriate for the case 'Literature Atlas'. However, the data experts judgment is different. The data expert of the case 'Brienz' shows little interest in using the visualisation for data analysis. The data expert of the case 'Literature Atlas' shows the greatest interest and the data expert of case 'Deer SNP' shows great interest in using the visualisation for data analysis. Thus, the proposition is rejected as no explicit relationship is existent. Other aspects which might influence the perceived appropriateness of the visualisations by the data experts are analysed in III.III below. Generally, all data experts state that it is useful to compare their data sets to the landscape in the 3D visualisation. Also their suitability ratings (section 3.5.4) of the visualisations are generally high.

✳ P III.II: A visual combination of data and landscape allows more efficient and/or more effective finding of insight into the data.

This proposition cannot be evaluated and answered directly from the collected data set. The data expert of the case 'Brienz' judges himself to be too familiar with the data set to see something new in it. All details of the data set, the circumstances of the collection and context information are well known. For example, a point below ”Äsch” shows a large positional shift in 2005. However, the movement was only in that year and area as just below the point a surface landslide happened and the area above (containing the measured point) moved afterwards. Also the Switzerland data set of the case ’Literature Atlas’ had already been analysed by the interviewed data expert and she remarks that nothing new can be found. The northern Germany data set is not yet finally analysed but the time spent with the visualisations is not long enough to find new insights. Additionally, the data expert is more interested in using the time exploring the possibilities of the visualisation rather than trying to (re-)evaluate the data set. However, the data expert from the case ’Literature Atlas’ states that she would like to see an new/unknown data set in this type of visualisation. She thinks it possible to find new insights when using such a display. In the case ’Deer SNP’ the data was analysed some time ago and looking at the data set got the data expert thinking about it anew. Two statements in table 3.4 show that new insights into the data might be found by using this type of visualisation. Additionally, in the case ’Deer SNP’ no current or not already analysed research question is currently of interest. Thus, the proposition can neither be accepted nor rejected on the base of the collected data. There are some hints that the visualisations may be efficient and/or effective but a direct evaluation of these properties are not possible. However, while the data experts are not able to see something new in their data sets they are able to show and explain their data sets and earlier findings with the help of the visualisations, thus, indirectly validating the visualisation technique through confirmation of what is already known.

✔ Q III.III: Can differences in appropriateness of the visualisation for data analysis be related back to different characteristics of the evaluated application settings?

Previous visualisation usage: The importance of the data-landscape relationship was proposed to directly influence the perceived appropriateness of the displays of quantitative information within virtual environments for the analysis of the case data sets. However, no evidence supporting this proposition is found (P III.I above) where positive statements of usefulness are not congruent with the data experts’ interest in using the 3D visualisation for data analysis tasks. However, the perceived and reported appropriateness of the tested visualisations for data analysis do co-vary positively together with the type of analysis previously done by the data experts. Especially in the case ’Literature Atlas’, where the data-landscape relationship seems less important than in the other two cases, the data experts mainly use visualisations for the analysis of their data and also report the type of visualisation tested here as useful for data analysis. In the case ’Brienz’ the data is mainly analysed numerically and visualisations are used for plausibility verification of results and regarded rather a tool for communicating findings to a wider public. The data expert is reluctant regarding the use of 3D visualisations for analysis tasks but would like to try its use for the communication of results to lay users. In the case ’Deer SNP’ visualisations (overlays of ortho imagery with data in GIS) were created to analyse the data and also to communicate the findings to a wider public. Especially, as the analysis of the data in regard to the landform or altitude is reported as being quite complex in GIS, the visualisations tested in this research are regarded as having some potential for the future analysis of data sets but also for communication purposes. Additionally, it is mentioned that this type of visualisation might be useful for comparison with data in different national parks. Data experts of other national parks, who know the type of data but not the environment the data relates to, may like to work with it. Thus, the data experts mainly want to use the ’new’ visualisation in the same way they used other visualisations before. This impression is strengthened by the data experts of the cases 'Brienz' and 'Literature Atlas' reporting that there is some effort needed to visually interpolate the point data to a data surface. Both of them have already used visualisations of interpolated surfaces for analysing (case 'Literature Atlas') or communicating (case 'Brienz') the data set.

Visualisation completeness: In all three cases it is remarked that some parts or aspects of the case study data sets are not possible to visualise in the tested visualisation type. In the case 'Brienz' the very important bearings of the slope movement vectors, which should be analysed together with the amount of movement, are not displayed. This may add essentially to perceiving the visualisation as less appropriate. In the other two cases 'Literature Atlas' and 'Deer SNP' the data experts also remarked about missing data. But the data missing in the case 'Deer SNP' is 'nice to have' data which was also not displayed in visualisations used so far. Similarly in the case 'Literature Atlas'. The missing path or areal data is interesting and important to analyse but the point data can also be analysed and makes sense without these data.

3.5.4. Further analysis of the case data

Visualisation suitability: The data experts from the cases 'Literature Atlas' and 'Deer SNP' like the visualisations and report that the interpretation of the landscape, height differences or slope steepness (case 'Deer SNP') is quite easy. Generally, it is stated that this type of visualisation gives a good overview of the data set and area where the data is located in. The data expert from case 'Brienz' is less enthusiastic but still thinks the visualisations helpful for comparing the data with the 3D landscape and the land-cover (ortho imagery).

Table 3.6 shows that all three data experts rated the visualisations somewhat suitable. Additionally, they like the visualisations and would, to slightly varying degrees, like to reuse this type of visualisation for similar data sets (even though they think the visualisations are not perfectly suitable). The slightly varying suitability ratings reflect the perceived appropriateness of the visualisations as evaluated in section 3.5.3.

The reference frames of the data graphics are noted as not being intuitive and crowding the display (case 'Brienz') and as being helpful (case 'Literature Atlas'). Bar charts displaying more than three or four variables are thought too complex and occluding, especially if the bars are arranged beside each other, to be useful (cases 'Literature Atlas' and 'Deer SNP'). Generally, data aggregation results in loss of detail while aiming for improved clarity of the display not only for the data graphics as tested in this study but also for other display types (e.g. statistical surfaces or classification for choropleth mapping (Slocum et al. 2005)). Beside the thematic aggregation also the spatial aggregation of the data sets is remarked upon by the data experts as potentially being a problem. For example, in the case 'Literature Atlas' some of the original data was already spatially aggregated. Additionally, different spatial aggregation might lead to different data graphics and thus potentially different interpretations. In case 'Deer SNP' the aggregation to the smaller grid is preferred as it results in less data per data graphic and thus less complex data graphics. Consequentially, in case 'Deer SNP' more interaction possibilities are requested, for example, split the data to different simpler displays and allow switching them on and off.

Navigation and interaction in Google Earth: The data experts of the cases 'Literature Atlas' and 'Deer SNP' commented positively about the navigational and interaction functionality of Google Earth. In the case 'Literature Atlas' with its two larger areas covered by data it was helpful to switch on the place name labels for orientation. Additionally, the oblique view from some distance was useful for getting an overview of the data set and then zooming in for seeing the details. The data expert reported that the navigational functionality was almost as useful as seeing the data as bars in the landscape. Similar statements were made by the case 'Deer SNP' data expert. She found the function of zooming in to see the details, such as the data distribution on a single slope, and especially zooming out to see an overview of the data and area very helpful. Additionally, she mentioned that it is very helpful to change the viewpoint, for example, not to look at the data in the usual south-north view but rather look, for example, from the top of a valley down along the valley to see the data distribution on both valley slopes. Provided some familiarisation with the navigational functionality it is even possible to look at single slopes from the side and get a very different impression of slope steepness and data distribution. The navigation is also very important to overcome data occlusions which may occur in data dense areas or in areas where the data is distributed over several valleys. The data expert from the case 'Brienz' thought the navigation of Google Earth slightly difficult and not intuitive. He also mentioned that some zooming behaviour had changed from the previous to the current version (Google Earth version 5.1.3533.1731 was used). However, after some familiarisation with the navigation it is quite useful and can give different views of the area and data. He liked the Google Earth functionality which allows clicking on data graphics in data dense areas and thus moving them temporarily apart to overcome data graphic overlaps.

Functionality like this might also be available in other tools. However, common (2D-) GIS are normally restricted to a south-north oriented viewing of data. This makes sense when working with traditional map-like displays which include labelling. The reading of labels in a rotated map can be difficult (as remarked by participants of the stage IIa experiment, section 3.6.3). The south-north view is often abandoned in most 3D views, for example in 3D block views or fly-overs. As the data experts remark, more freedom in viewpoint choice is potentially useful. Additionally, as the spatial arrangement of geo data is predefined (except for different spatial aggregations) we may gain from looking at such data sets from unusual perspectives.

3.6. Analysis of participant s comments in all research stages

3.6.1. Quantitative data graphics

Background frames: The number of positive comments indicating the helpfulness of background or reference frames is very high in stage I and decreases slightly in stage IIa (figure 3.54). In stage IIb only negative comments regarding the background frames are noted. The main critique being that the frames overcrowd the display and that they are confusing and disturbing. The background frames seem to become less helpful the more complex the data is which is displayed in the 3D virtual environment and the more complex the tasks are which are to be fulfilled. On the other hand, background frames are needed to show which data belongs together and where a data graphic belongs (figure 3.55). In the case studies a positive and a negative comment regarding the background frames is reported. Thus, the views of the data experts (table 3.5) regarding the usefulness of background frames vary. However, they did not actually work with the data displays and answer research questions (section 3.5.3, proposition III.II). In addition to the background frames, a participant in stage IIb and a data expert of stage III asked for a legend or scale allowing judgements of absolute lengths of the bars.

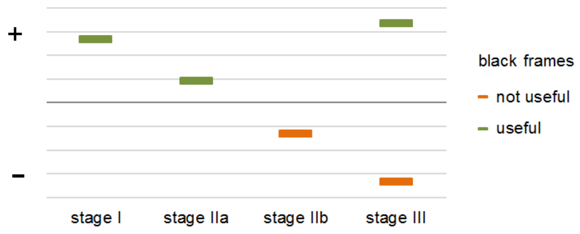


Figure 3.54.: Comparison of positive and negative comments regarding the background frames in all research stages (relative to the number of participant's, table 2.1)

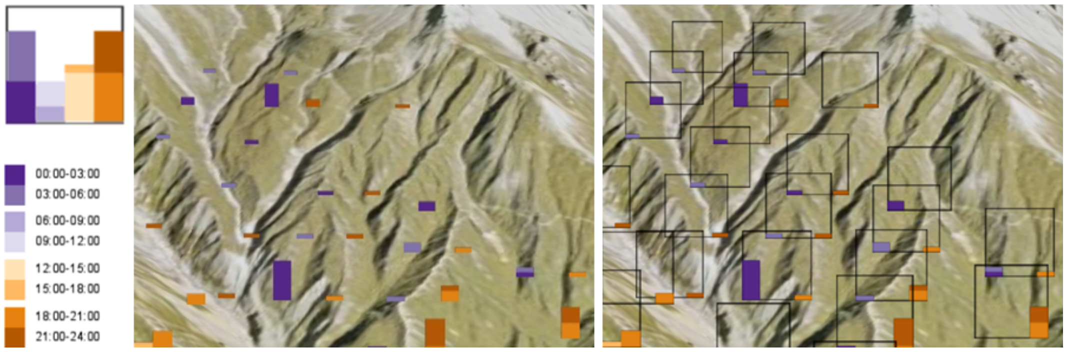


Figure 3.55.: Example of stage IIb data graphics without and with background frames)

Data density: In stage I no participant commented on data density while in stage IIa occasional comments on overlapping data graphics and landscape features occluding data graphics are noted. In 3D, most of these statements are followed by the remark, that the possibility to navigate makes this a small problem. In the 2D visualisations of stage IIa some data graphics were slightly displaced to avoid symbol overlaps. However, with more dense data sets, as for example in stage IIb, this is no longer possible in all cases and a different solution for avoiding data graphic overlaps is needed. Thus, in stage IIb the data graphics were available in two different sizes in the 2D visualisations making it possible to avoid data graphic overlaps by choosing the smaller size. Additionally, the data graphics always occlude map symbology in 2D but this is not commented upon. Shepherd (2008) reports that, especially for point symbols, overlap in 2D is a largely unrecognised problem. In 3D it is possible to circumvent map symbology overlaps through data graphics by navigating to a different viewpoint. In the larger data sets of case study 'Deer SNP' the data expert noted that navigation is a natural solution to overcome data occlusion, be it overlapping data graphics or data occluded by mountain ridges. Especially, zooming out to get an overview and zooming back in to areas of interest is helpful for avoiding data occlusion and for data analysis in general.

In stages IIb and III comments regarding data graphic complexity show up. The problem of occlusion and data graphic overlaps is still mentioned but additionally, participants comment on the impression that if more than 3-4 variables are displayed in a bar chart then such data graphics are too complex for interpretation. Single bars or stacked bars displaying a maximum of four variables are preferred over more complex charts with stacked bars and/or bars beside each other. This also mitigates the problem with data graphic and/or symbology overlap and occlusion.

Relation between data graphics and landscape: In stages IIa and IIb several participants noted that it is difficult to judge the exact point or area that a data graphic relates to. Based on the number of comments this is perceived as being most problematic in IIa in 2D and a little less in stage IIb in 2D. In the 2D representation it is unclear if the data graphic relates to the point at the bars base or if it relates to the area a bar covers. Some note the same problem of exactly relating the data graphics to the landscape for stage IIa and IIb in the 3D virtual environments. However, for 3D it is also remarked that navigating around a data graphic makes the area it relates to visible.

3.6.2. Tasks

Comments on tasks are mainly given in stage IIa where seven tasks had to be answered. Figure 3.56 shows the comparison of the relative number of comments stating that a certain task is difficult to fulfil or not understood (the data is not corrected for participants stating this several times during the same stage for the same or different tasks). Most comments about difficult to understand tasks contain hints that participants did not read the introduction to the experiment carefully (e.g. ”what do you mean with bars?”). Others had difficulties understanding words like ”area” or ”location”. Comments about difficult to understand tasks are only visible in stage IIa were participation took place via an online questionnaire and participants were not able to ask questions. In stage I and IIb participation took place in a face-to-face setting. Only few questions needed answering during the face-to-face experimental sessions (though the questions were of similar nature, e.g. ”what do the bars exactly mean”, ”what is meant by area”). Additionally, the tasks in stage IIa are reported as being difficult to answer. It is the case that some tasks are easier to answer with certain stage IIa data sets and difficult or with no clear answer in other data sets. For example, task t5 (”In which area(s) does the number of deer visits increase with increasing altitude?”) was answered with small scale variations or statements that no such pattern can be found in six of the eight data sets (W1, W2, W3, W4, S3 and S4). These ’no’ answers were evaluated like the other insights by rating them for complexity (’low’ without additional information) and plausibility (dependent on the data set). The more open task of stage IIb which left the participants more freedom in the exploration of the data set by requiring them to report insights based on one general question led to less remarks about task difficulty. Several participants noted in stages IIa and IIb that finding patterns and comparing data to location and landform needs considerable effort. A few participants of stage IIa also state that such visualisations (in 2D and 3D) can only be sensibly used for getting an overview of a data set and not for finding more detailed insights.

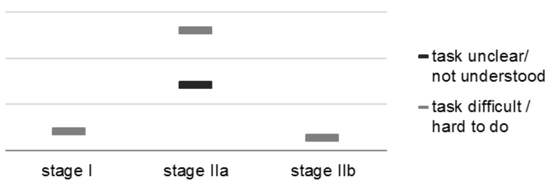


Figure 3.56.: Quantification of comments stating that a task is difficult to fulfil or not understood (multiple comments by one participant possible, relative to the number of participant’s, table 2.1)

3.6.3. Visualisations, interaction and navigation

In all stages participants asked for different visualisations or different visualisation contents. In stage I, one comment asks for reference objects such as trees or buildings for easier size judgement. In stage IIa the comments are more numerous asking for more context, details on demand, different drapes or layers and also imagery with better resolution. Additionally, they would like to switch on/off layers and also compare data sets by interaction (e.g. selecting which (sub-)data sets are visible).

Other visualisations they ask for are density surfaces or heat maps, slope and curvature maps and dot maps (not aggregated data). According to them it should be possible to switch between different visualisations and also make more details available such as exact numbers, numeric distribution or linked graphs. In stage IIb participants mainly asked for additional data in their comments. For example, several of them would like to see the deer routes and/or more detailed date and time information to explore the deer's movement patterns. This is information the data expert of case 'Deer SNP' would also like to see displayed. The data experts of stage III also ask for data displays (e.g. density surfaces) they have already used for their data set (section 3.5.3, question III.III). A few comments in all stages ask for the improvement of the legend (more detail) and that a scale bar should be displayed in the 2D and 3D visualisation. Additionally, a few participants remark on the rendering quality of the imagery and map background in 2D and especially in 3D. The quality of the background maps and ortho imagery was the same for 2D and 3D (section 2.6 for an example of rendering quality). However, in the 3D visualisation map symbology, especially labels of spot heights, may depending on the current viewpoint, be difficult to read. Several participants remarked upon these difficult to read spot height labels from the map in the 3D visualisation.

The above statements already include some comments regarding interaction and navigation. Additionally, a participant in stage I noted that zooming in and rotating the scene helps the judgement of the bar sizes. In stage IIa several participants noted that clicking the map layer in Google Earth on and off and looking at the ortho imagery provided by Google Earth helped them even though this functionality was not intended to be used. Additionally, two participants noted that they navigated in 3D to a top view to see the problem as a 2D map. Case study data experts noted that several visualisations of data subsets with easier data graphics might be integrated through an interface where it is possible to switch between them. In stage IIb the participants could switch between different backgrounds (map or ortho imagery) for the data analysis and some commented about this positively.

The general comments about the usefulness or difficulty of navigating or interacting with the 2D and 3D visualisations are numerically shown in figure 3.57. Especially in stage IIa in 3D, interaction and navigation is thought essential for fulfilling the tasks. Additionally, the navigation and/or interaction in seem similarly unfamiliar or difficult in 2D (interactive SVG) and in 3D (Google Earth).

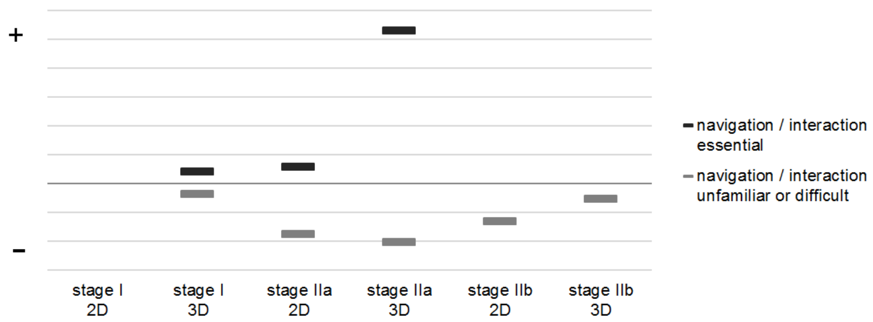


Figure 3.57.: Comparison of relative numbers of comments regarding interaction and navigation in 2D and 3D (relative to the number of participant's, table 2.1)

3.6.4. Suitability

Analysing the comments from stages IIa and IIb (only one comment in stage I; suitability ratings of the case study data experts of stage III in table 3.6) details the visualisation preferences by the participants. Figure 3.58 shows that in the comments for stage IIa and IIb about the same number of statements saying that the 3D visualisation is better can be found. In stage IIa there are also several comments rating the 2D visualisation better or making no difference between 2D and 3D. The number of comments saying that 2D is better suited for the tasks at hand is lower in stage IIb. In stage III the data experts rated the suitability of the visualisations quite positively (section 3.5.4).

Figure 3.59 shows the number of comments regarding the efficiency of the 2D and 3D visualisations in general and for judging landform/altitude and location/land-cover. The numbers of comments vary but generally, the 3D visualisation is thought to help the judgment of landform and altitude more than the 2D visualisation. Looking at the comments, several statements are found which defend the participants' map reading skills thus preferring the 2D visualisation. They would not like to see 2D visualisations replaced by 3D visualisations. They also remark that potentially 3D visualisations are for people who are not as good at map reading. In contrast to the statements preferring the 2D visualisation, there are several participants making statements like "I was surprised at how good GE was for enabling me to see the nature of the variation in topography. The 3D map made it easy to find the highest points in the landscape and less guessing about the landscapes fluctuations between altitude lines need to be implied." or saying that the interpretation of the height information from the 2D map needs effort and that it is easy to overlook height or landform related information in 2D. There are also a number of participants who commented positively or negatively about both 2D and 3D visualisation and visualisation aspects.

Few participants comment that efficiency might not be an issue with these types of visualisations as complex data analysis needs time and that the time should be taken to get accustomed to the visualisation type and familiarise with the data set.

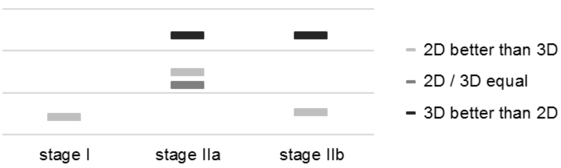


Figure 3.58.: Quantification of comments stating a preference for either 2D, 3D or that the two types of visualisations are equal (relative to the number of participant's, table 2.1)

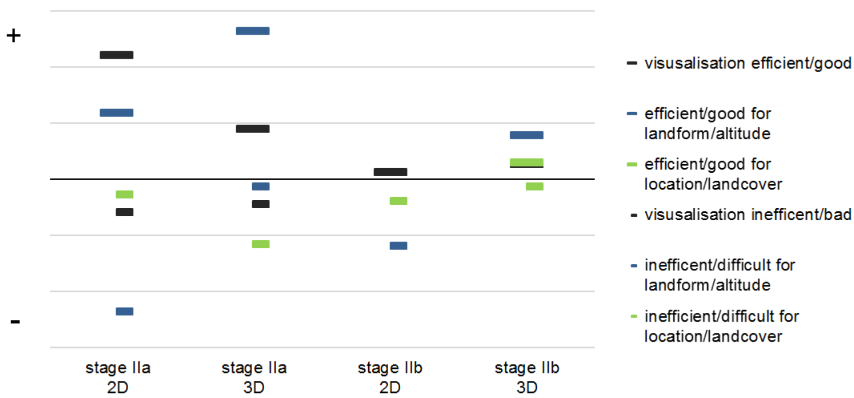


Figure 3.59.: Quantification of comments regarding the efficiency of 2D/3D visualisations in general and for judging landform/altitude and location/land-cover in stages IIa and IIb, no comments in stage I, few comments in stage III (efficiency = positive values, inefficiency = negative values, multiple comments possible, relative to the number of participant's, table 2.1)

Evaluating performance of participants stating that either the 2D or the 3D visualisation is better

For stage IIa the efficiency and effectiveness of task performance is evaluated for participants stating explicitly that either the 2D or the 3D visualisation is better suited for the tasks (figure 3.58). As one participant stated that both visualisation types are better and the participants stating that the 2D visualisation is better stated this several times, finally, the performance for three participants stating that the 2D visualisation and seven participants stating that the 3D visualisation is better is compared. Table 3.7 shows data graphics for the different measures. Statistical significance testing of the difference is not done because of the small numbers but trends can be identified. Summarising the information in table 3.7, the '3D better' participants found about 1.5 times the number of insights in 2D and 3D than the '2D better' participants. Both groups found slightly more insights in 3D. The same pattern is shown in the task performance times. Finding and reporting more insights also needs more time. Confidence ratings are interesting for the '2D better' group. Contrary to the hypothesis that they might be more confident in 2D if they think 2D is better, they report more insights of high plausibility and less insights of low plausibility in the 3D visualisation. Similarly for the complexity of insights. The '2D better' group reports only insights of high complexity in the 3D visualisation. Also the '3D better' participants report more insights of high complexity in 3D. For plausibility, both groups report more highly plausible insights in 2D which may be surprising for the '3D better' participants. A clear difference between the two groups is visible in the use of different references (location L, altitude A and both LA) in the reported insights. The '2D better' participants report a high number of insights with altitude references in 2D and 3D. They may want to show that they can read the altitude and altitude variation from the map thus specifically report such insights. The '3D better' participants make more use of combined location and altitude references and report fewest insights with location reference in 3D.

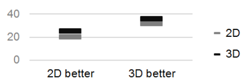
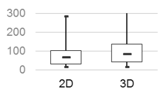
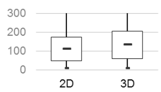
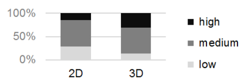
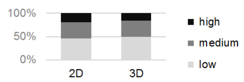
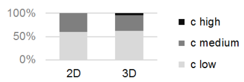
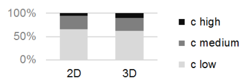
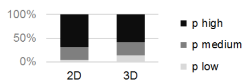
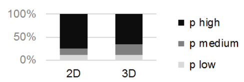
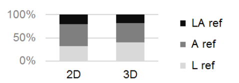
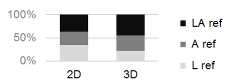
measure	2D better than 3D (3 participants, data sets W2, W3, S2, S4)	3D better than 2D (7 participants, all data sets)
Ø number of insights per participant		
time [seconds]		
confidence ratings		
complexity		
plausibility		
location / altitude reference		

Table 3.7.: Comparison of different performance measures for participants stating that either the 2D or the 3D visualisation is better (time: left graphic, maximum value 3D=465; right graphic, maximum values 2D=537, 3D=830)

4. Discussion

< Chapter 3. Data analysis and results

The results of the research stages I, IIa, IIb and III which use increasingly complex data and tasks to evaluate the appropriateness of quantitative data displays in virtual environments while moving from an 'in vitro' experimental setting in stage I to 'in vivo' applied case studies in stage III are thematically structured and discussed across all stages in this chapter.
2D bars and bar charts are appropriate for displaying quantitative data in virtual environments (4.1.1). They should not show more than 3-4 variables and the reference frames are less helpful when the data is more dense as they crowd the display. Two single bars can be compared but the length of the smaller bar is slightly overestimated (4.1.2). Navigation is important to overcome data graphic and landscape occlusion and also for data analysis as it allows viewing the data and/or topography from various viewpoints (4.1.1 and 4.1.3.1). Comparing the 2D and the 3D visualisations generally through task completion times, user confidence, complexity and plausibility of the evaluated insights shows no significant differences between the two display types (4.1.3.1). The 3D virtual environment does not foster a more efficient or effective analysis of the data and landform relationship as compared to the 2D representations (4.1.3.2). The participants' preferences for either the 2D or the 3D visualisation cannot completely be evaluated. However, participants preferring the 3D visualisation tend to take more time and report more insights (4.1.3.3). The word usage differences between 2D and 3D in the insights point to the problem of reading spot height labels in potentially rotated views in the virtual environment (4.1.3.2).
The tasks along the research stages I, IIa and IIb are of increasing complexity. A functional view of data and tasks allows tasks referring either to location, altitude or a combination of location and altitude to be separated (4.1.4). The reference type and the wording of tasks strongly influences the task answers as found in stage IIa. Thus, in stage IIb a single synoptic task is employed. Tasks referring to a combined reference set of location and altitude are solved less confidently.
Another strong influence on the results is found in the data sets and settings (4.1.5). While for geodata the setting (the area a data set belongs to) is normally not separable from the data, such a separation was simulated in this research by spatially transforming some data sets to another area. The two settings W and S and the corresponding data sets in stage IIa and IIb show some differences in land-cover, spread of data points and topography. The eight data sets of stage IIa are analysed to tentatively identify some characteristics which ask for visual analysis of the data set in either the 2D or 3D visualisation or both. The proposition is strengthened by data experts statements and observations from stage III. Generally, small bars are difficult to locate and compare.
Comparing the results of stage IIa and IIb allows the influence of increasing data and task complexity, as well as the change in methods by switching from answering seven tasks to reporting insights based on one synoptic task, to be closely observed (4.1.6). Participants of stage IIb reported insights with higher complexity and confidence, slightly less plausibility and potentially under less time pressure compared to stage IIa. Importantly, they reported many more insights referring to location in stage IIb. Combining this finding with statements from stage III and the reported task difficulties in stage IIa it seems that participants are not readily used to analysing data in relation to altitude and landform. Further findings include the suitability of either 2D and 3D visualisations, the influence of previous visualisation usage by the participants and that the visualisations in applied settings need to show all important aspects of a data set (4.1.7).

The second part of this chapter reflects upon the methodological framework and the methods employed along the research stages from 'in vitro' to 'in vivo' (4.2). First, the methods and results of each research stage are summarised (4.2.1). Stage I employed a controlled experimental setting testing two elementary tasks in 2D and 3D. The results allowed the proposed hypotheses to be tested and concluded that 2D bars on billboards can be compared. Stage IIa used an experimental setting where participants answered seven tasks in two 2D and two 3D visualisations. Analysing the data collected allowed all hypotheses about 3D supporting the analysis of data in relation to landform better than 2D representations to be rejected. Stage IIb employed an experimental setting where participants reported insights in a more complex data set. The results allowed all hypotheses about 3D supporting the analysis of data in relation to landform better than 2D representations to be tested. Stage III explored if quantitative bar displays in virtual environments are appropriate for applied scenarios by using a case study approach. While valuable results could be gathered from the data collected it was not possible to evaluate if data experts gain more efficient and effective insight into their data in relation to altitude as the data experts did not work with the data set as anticipated. The typical research questions reported in stage III are comparable to the type of questions used in the earlier stages of the research.
A sequential mixed methods approach employing four research stages was designed and implemented in this research (4.2.2). This allows conclusions to be drawn forward and backward along the research stages. While in the forward process findings from an earlier stage inform later research stages backward inferences are done by revisiting and re-evaluating earlier collected data and findings in light of later findings. It is described how the combination of methods along the bridge complement, expand or triangulate (the reasons for conducting a mixed methods approach) the findings exemplified by the evaluated aspects data graphics, participants, tasks, data sets and context information. Additionally, employing balanced within-subject designs in the experimental stages is sensible (4.2.3). The selected appropriateness measures are useful even though the answers needed splitting into insights. The issue of implementing either online or face-to-face settings may need consideration in future evaluations (4.2.3). The chapter is concluded with a final reflection on bridging methodologically between 'in vitro' and 'in vivo' research designs (4.2.4). The concept of overcoming weaknesses of one method through strengths of another method proved useful. It was detected that real world applied settings may not always be most complex in data and tasks even though they include most context information. Additionally, the bridge could be extended further on both the 'in vitro' and especially also the 'in vivo' end. Implementing longitudinal case studies may yield more information about the use of a visualisation type 'in vivo'. Evaluating the visualisation through a mixed methods approach may be less profound than a study researching a single aspect in-depth. However, it is more holistic and may yield applicable results faster. This is especially important as future technological developments will not wait for theory about appropriate visualisations and applications to be developed first.

Chapter 5. Conclusions and outlook >

4.1. Summary and discussion of results and findings

The following sections discuss the results and findings detailed in chapter 3 in light of various aspects and measures integrative over all research stages where the discussed topic is evaluated (references back to the relevant sections of chapter 3 are included to allow looking up the details of statistical testing and the graphical displays of the discussed results). The corresponding research hypotheses (section 1.4) are listed in the respective sections and acceptance or rejection of the hypotheses indicated. The characteristics of each research stage is described in chapter 2 and table 2.1 offers a quick overview of all stages. Summarising, the research stages I, IIa, IIb and III use increasingly complex data and tasks to evaluate the appropriateness of quantitative data displays in virtual environments while moving from an 'in vitro' experimental setting in stage I to 'in vivo' applied case studies in stage III.

4.1.1. Quantitative data graphics

The use of 2D bars and bar charts (combination of single bars) on billboards in the virtual environments is derived from 2D cartography and knowledge about the human perception (section 2.3). The examples of data representations in virtual environments (figure 1.3) show that other displays of quantitative data are possible. But taking into the account the potential distortions of data displays through depth cues such as size gradient and the opportunities to navigate and, thus, potentially looking at the data representations from unfortunate angles, leaves the 2D bars on billboards as the most promising option. A different study (Bleisch accepted) evaluates six quantitative data displays in virtual environments (2D bars, 2D circles and 3D bars each of them with and without reference frames) with two simple tasks as in stage I of this research. The results show that effectiveness is best for 2D bars with frames, then 3D bars with frames, 2D bars and 3D bars without frames in this order decreasing with circles with and without references frames being worst (Bleisch accepted). In 2D information visualisation it is generally not advisable to use multidimensional symbols if a lower-dimensional symbol would communicate the same fact (Siegrist 1996). However, the authors own observations and a comment noted during stage III of this research suggest that the 2D data graphics on billboards do not integrate too well into the otherwise three-dimensional environment and 3D bars may be more suitable (figure 1.3h for an example of 3D bars in a virtual environment). Such aesthetic considerations can affect the usability of visualisations as shown by Cawthon & Vande Moere (2007). As 2D and 3D bars perform similarly well for elementary tasks (Bleisch accepted), future evaluations of data visualisations within virtual environments could test 3D bars and 3D bar charts with more complex tasks and data sets similarly as is done in this research with 2D bars and bar charts on billboards. Additionally, it may be useful to evaluate other types of data displays potentially suitable for virtual environments (e.g. Kumke 2009, use of cones to display data values in 3D city models).

The data graphics (section 2.3.1) employed in the four stages of this research become more complex from stage I to stage III according to the increasing complexity of the data sets and tasks. In stage I two and in stage IIa several single bars representing univariate data are evaluated. In stage IIb and III bar charts (combinations of several single bars stacked and/or beside each other representing multivariate data) are used.

In stage I two bars were compared in 2D and 3D (section 2.5.1). In 3D the bars with reference frame resulted in faster task performance and users being more confident in their findings (table 4.1, Bleisch et al. (2008)). In 2D no significant difference in efficiency is found. Based on this, bars with frames are used in 3D settings and bars without frames in 2D settings (avoiding the use of display space for frames without additional benefit) in all subsequent research stages. However, as analysed in section 3.6.1 and shown in figure 3.54 the background frames are becoming less helpful when the data graphics are more complex (bar charts) and the display more dense (e.g. in overview situations when the data graphics overlap) as they clutter up the view. However, in views showing details the background frames are needed to group the data together and locate it.

		stage I	(Bleisch et al. 2008, ➪ Appendix A)
hypotheses	✘ ✔	H I.III: The efficiency of task completion is improved by the use of a reference grid (frame) (in the 2D ✘ and 3D ✔ settings).	

Table 4.1.: Tested hypothesis about the efficiency of reference frames (✔ = accepted, ✘ = rejected)

Data density: Comments in stages IIa and IIb report that overlaps and occlusion are a problem (section 3.6.1). The problem of hidden objects, overlaps and occlusion in 3D virtual environments is a recognised and researched problem (Zhu & Chen 2005, Tory et al. 2006, Elmqvist & Tudoreanu 2007, Elmqvist & Tsigas 2008). In this research, data graphics occlude map symbology especially in the 2D representations but overlapping data graphics are also a problem in the 3D visualisation. However, in the 3D visualisations navigation is mentioned as a natural solution to overcome data overlap and occlusion (confirming Shepherd's (2008) proposition). Especially, zooming out to get an overview and zooming back in to areas of interest is helpful to avoid data occlusion and for data analysis in general. This corresponds to the first part of Shneiderman's (1996, p. 2) information seeking mantra; "overview first, zoom and filter, then details-on-demand". Additionally, the navigation as implemented in Google Earth supports this behaviour as it is designed to zoom in and out (or change the level of detail) rather than to move/fly along the earths surface. The navigation and interaction possibilities offered by Google Earth are mainly positively commented upon by stage IIa, IIb participants (section 3.6.3) and stage III data experts (section 3.5.4). They like the freedom of viewpoint choice and, thus, being able to look at the data and/or topography from various viewpoints. Nevertheless, a number of comments (section 3.57) remark on the difficult and unfamiliar navigation in 2D (SVG) and 3D (Google Earth).

From the participants' comments in stage IIb and III a complexity threshold for data graphics is established (section 3.6.1). While in stage IIb potentially eight variables are shown in one data graphic most data graphics show four or less variables. But participants prefer single bars or stacked bars displaying a maximum of four variables over more complex charts with stacked bars and/or bars beside each other. Such simpler displays could potentially (as asked for by some participants, section 3.6.3) be interactively connected to compare different (sub-) data sets side-by-side. Brath (1997, p. 109) recommends using "the lowest number of dimensions which solves the task". Principles for glyph design in 2D visualisations are explained by Ward (2008). He admits that the number of data values that can be effectively visualised is constrained but does not advise on a maximum number of values but on using glyph displays only for "modest-sized data sets" (Ward 2008, p. 180). In this research, in addition to data graphic complexity, it is noted that especially in 2D it is difficult to know where a bar relates to (e.g. to the area at the bottom of a bar or the area a bar covers). The same problem exists in the 3D visualisation as well but there it can be overcome by navigation (section 3.6.1).

4.1.2. Accuracy of data graphic interpretation

The accuracy of data graphic interpretation is tested in stage I (section 3.1). For testing accuracy, tasks with exact answers are essential. The tasks of stage I, finding the higher bar and comparing the lengths of two bars, lead to such exact answers. The results show that the taller bar could be identified in almost 100% of the cases in 2D and 3D. When the smaller bar was sized 86% or more of the larger bar participants often judged the two bars as being of equal size. Also the comparison of the two bars is done successfully. The mean of the differences between estimated and exact comparison is 1.68 in 2D (without reference frames) and 1.64 in 3D (with reference frames). Thus, the participants tended to slightly, but significantly, overestimate the size of the smaller bar compared to the size of the taller bar. Table 4.2 gives and overview of the tested hypotheses based on the evaluation results (Bleisch et al. 2008, ➩ Appendix A). The evaluations of all following stages are based on these results claiming that the higher of two bars can be identified and that two bars can be compared in 3D with the same accuracy as in 2D.

		stage I	(Bleisch et al. 2008, ➪ Appendix A)
hypotheses	✔	H I.I: Users are able to identify the taller of two bars in a 3D desktop virtual environment as well as they can in static 2D graphics.	
	✔	H I.II: The effectiveness of estimating differences between two bars is not significantly different in the 2D and 3D settings (2D without frame, 3D with frame).	

Table 4.2.: Overview of tested hypotheses about basic bar comparison in 2D and 3D (✔ = accepted, ✘ = rejected)

4.1.3. Comparing 2D and 3D visualisations

4.1.3.1. Overview

Benchmarking the 3D visualisation against a 2D representation (section 2.1) is based on the fact that 2D representations are traditionally used for data and geodata analysis and, thus, thought to be more familiar. Additionally, the design of the data graphics in 3D was derived from literature about cartographic and perception knowledge which was gained by studying 2D data representations (section 2.3.1). While the 2D representations based on the principle of comparing data to a map may be more familiar, participants do not seem to be familiar with interactive 2D representations. Several comments (section 3.6.3 and figure 3.57) remark on the unfamiliar or difficult navigation and interaction functionalities in both the 2D and 3D visualisations. The navigation/interaction with the 2D visualisations (interactive SVG displays in stage IIa and IIb) is similarly often remarked as being unfamiliar or difficult as is the navigation/interaction with the 3D virtual environment in Google Earth. However, it is mentioned considerably more often that navigation/interaction with the 3D virtual environment is essential for the interpretation of the data, for example to overcome data graphic and/or landscape overlaps, to relate the data to landform, to get a good view on slopes or other topographic features or to look at the data from a different than the usual south-north oriented view (sections 3.5.4, 3.6.3 and figure 3.57). The latter is interesting as earlier studies (Cerny & Wilson 1976, Edsall & Deitrick 2009) have shown that orientation is an important factor for data recognition and analysis. The spatial arrangement of geodata is predefined through the three-dimensional spatial reference. Thus, re-expression (Tukey 1977, DiBiase, MacEachren, Krygier & Reeves 1992) possibilities of the spatial data component are very limited. Viewpoint changes may nevertheless help to get a different impression of the same data even though this may be a mixed blessing. While a different impression may support data analysis it may impede it as well. Navigation and interaction in 2D and especially 3D displays is generally regarded as a (cognitive) cost (e.g. Wickens & Baker 1995, Ware & Plumlee 2005, Nielsen 2007, Shepherd 2008) but it is important, as is remarked upon by the participants of this research and reported in a slightly different context (Ball & North 2008). Sebrechts et al. (1999) report that these costs decrease with increasing experience with 3D displays. In addition to the experience the informed participants of this research (section 2.4) already have, they were given time to familiarise with the 2D and 3D visualisations in each research stage (section 2.5). Another viewpoint is reported by Keehner et al. (2008) who conclude that visibility of task-relevant information is more important than interaction with the 3D visualisation.

Task completion times: Comparing bar lengths in stage I took significantly longer in the 3D setting than in the 2D setting (Bleisch et al. 2008) (table 4.3). In the stages IIa (section 3.2.1.1, figure 3.7) and IIb (section 3.3.1.1, figure 3.28) no significant differences between answering tasks (IIa) and finding insights (IIb) are found between the 2D and 3D setting. The time difference between 2D and 3D in stage I most likely comes from the fact that in stage I the 2D display was static while the 3D display allowed navigation. Ware & Plumlee (2005) report that eye movement is a much faster type of navigation than zooming or even flying. In stages IIa and IIb both the 2D and 3D display allowed interaction and navigation. The case study data experts belong, regarding task completion time of the stage I tasks (finding the higher bar and comparing two bars), to the same group of informed users who also participated in stage I (section 3.5.1, figure 3.53).

		stage I (Bleisch et al. 2008, ➪ Appendix A)
hypothesis	✘	H I.IV: The efficiency of estimating absolute values from bar lengths (with a reference frame) is not significantly different in the 2D and 3D settings. (3D takes longer)

Table 4.3.: Overview of tested hypothesis about the efficiency of basic bar comparison in 2D and 3D (✔ = accepted, ✘ = rejected)

User con dence, complexity and plausibility of insights: Evaluating the user confidence, complexity and plausibility of insights (table 4.4) shows that there is no significant variation between 2D and 3D either in the results of stage IIa or of stage IIb. In detail the variation in the users confidence ratings between 2D and 3D is not significant either in stage IIa (section 3.2.1.2, figure 3.8) or in stage IIb (section 3.3.1.2, figure 3.29). The variation in the complexity of insights between 2D and 3D is not significant either in stage IIa (section 3.2.1.3, figure 3.9) or in stage IIb (section 3.3.1.3, figure 3.30). Also the variation in the plausibility of insights between 2D and 3D is not significant either in stage IIa (section 3.2.1.4, figure 3.10) or in stage IIb (section 3.3.1.4, figure 3.31).

		stage IIb (section 3.3.2.2)
hypothesis	✔	H IIb.II: The complexity and plausibility of insights reported do not vary significantly between the 2D representation and the 3D virtual environment.

Table 4.4.: Overview of tested hypothesis about complexity and plausibility of insights (✔ = accepted, ✘ = rejected)

Participants comments: Evaluating participants' comments about the suitability of the 2D and 3D visualisations (section 3.6.4 and figure 3.58) shows that some like 2D more while others prefer the 3D visualisation and a third group states that it does not matter if the visualisation is 2D or 3D. The evaluation of task performance in relation to participants' preferences was done as an example for stage IIa (table 3.7) hypothesising that participants may show better performance in the visualisation type they prefer or think more suitable. The results show that this is not the case but rather the contrary, for example, for confidence ratings or the complexity of insights where the participants who like 2D better perform better in 3D. However, this comparison is influenced by the fact that the participants who like 2D better did not work with all data sets (only data sets W2, W3, S2, S4). Thus, the variation between 2D and 3D visualisation may potentially be caused by the different data sets evaluated in 2D and 3D (section 4.1.5) as well as by the different participants and their preferences. The results for the participants who like 3D better are more homogenous between 2D and 3D (table 3.7). The only obvious difference between the two participant groups is that the participants who like 3D better reported in average about 1.5 times as many insights in 2D and 3D than the participants who like 2D better. Thus they spent more time finding and reporting insights. According to that visualisation type preference may rather effect motivation to work with a visualisation and spend time with it than result in better task performance. Instead of evaluating comments, further experiments in this area could explicitly ask for participants' preferences either before and/or after doing the experiment. This would allow a more systematic evaluation of potential effects resulting from participants' preferences.

4.1.3.2. Relation of data and landscape

Research hypotheses: Most of the research hypotheses and propositions of the stages IIa, IIb and III explicitly formulate ideas about the appropriateness of the data displays within virtual environments for the analysis of the data in relation to landform. Table 4.5 gives an overview of these hypotheses and indicates acceptance (✔) or rejection (✘) of them based on the evaluation results (chapter 3).

		stage IIa	(section 3.2.2)
hypotheses	✘	H IIa.Ia: Tasks where the reference set is location are solved more efficiently in the 2D representation.	
	✘	H IIa.Ib: Tasks where the reference set is location are solved more effectively in the 2D representation.	
	✘	H IIa.IIa: Tasks where the reference set is altitude are solved more efficiently in the 3D virtual environment.	
	✘	H IIa.IIb: Tasks where the reference set is altitude are solved more effectively in the 3D virtual environment.	
	✘	H IIa.IIIa: Tasks that include combined reference sets (location and altitude) are solved more efficiently in the 3D virtual environment.	
	✘	H IIa.IIIb: Tasks that include combined reference sets (location and altitude) are solved more effectively in the 3D virtual environment.	
		stage IIb	(section 3.3.2)
	✘	H IIb.Ia: Insights into the relationship between data and location/land-cover/etc. are more often and more efficiently reported in the 2D representation.	
	✘	H IIb.Ib: Insights into the relationship between data and altitude/landform are more often and more efficiently reported in the 3D virtual environment.	
		stage III	(section 3.5.3)
propositions	✘	P III.I: Displays of quantitative data within virtual environments are more appropriate for the visual analysis of the data if the data and landscape (2D/3D) relationship is more important in an application.	
	✳	P III.II: A visual combination of data and landscape allows more efficient and/or more effective finding of insight into the data.	

Table 4.5.: Overview of tested hypotheses and evaluated propositions about data and landform relationship analysis (✔ = accepted, ✘ = rejected, ✳ = not answered)

All hypotheses or propositions of the research stages IIa, IIb and III which assume that the relationship between data and landform or altitude is more efficiently or more effectively analysed in the 3D virtual environment, or that vice versa the relationship between data and location or land-cover is more efficiently or more effectively analysed in the 2D representation are rejected or cannot be answered (table 4.5). The 3D virtual environment does not foster a more efficient or effective analysis of the data and landform relationship as compared to the 2D representations. In stage IIa it is even found that the 2D representation yields more highly plausible and less low or medium plausible insights for tasks with a combined reference set location and altitude compared to the 3D visualisation (section 3.2.2.2, figure 3.15). In stage IIb no such trend is observed and the minimal variation is insignificant (section 3.3.2).

Wording of insights: In stages IIa and IIb the reported insights were analysed regarding their use of words belonging to different location or altitude related categories (section 2.6.2.4 for the word categories). The results show that the use of words is significantly different between 2D and 3D in both stages IIa (section 3.2.1.5, figure 3.11) and IIb (section 3.3.1.5, figure 3.32). In stage IIa L2 words (land-cover words such as forest, scree or grassy) and A3 words (form words such as steep, lower and highest) are more often used in 3D while A1 words (exact altitude values such as 1950m or 2400m) are used more often in 2D (figure 3.11). In stage IIb L1 words (exact grid values such as D3 or F4), A1 words (exact altitude values such as 1950m or 2400m) and A2 words (landform words such as mountain, slope or ridge) are more often used in 2D than in 3D (figure 3.32). Comparing these differences between stages IIa and IIb is difficult as participants reported more insights referring to location in stage IIb than in stage IIa (as discussed below in section 4.1.6). However, consistency is found for the lower use of A1 words (exact altitude values such as 1950m or 2400m) in the virtual environment. This is explained from a number of comments noting that exact height values (A1 words) are difficult to read from the map overlay in the potentially not north-oriented view of the 3D virtual environment (section 3.6.3). Comments on the grid labels (L1 words) do not show up. Grid labels are implemented as placemarks in Google Earth which always face the viewer and are thus readable independent of the viewpoint. The labelling of height values needs optimisation for 3D virtual environments. However, it was decided that displaying them as placemarks is not a solution as such height label placemarks would be too prominent and, thus, may compete with data value displays beside potentially causing occlusion of the important data displays. Harrower & Sheesley (2005) state that labels in virtual environments may serve as landmarks. Something that is labelled seems important. But this is an effect which is not wanted and thus lower readability of labels in the virtual environment was accepted in the design of the experiments. Other researchers (e.g. Lehmann, Trümper & Döllner 2011, Maas et al. 2007, Maas & Döllner 2006) have presented results on optimised labelling in 3D virtual environments but further research in labelling of combined displays of data and landscape may be needed where the labels are background information, for example embedded in a map, but are nevertheless important and so should be readable from various viewpoints equally well. Other possibilities would include interactive height querying, for example, a click onto the landscape would display the corresponding height value.

4.1.3.3. Summary

The differences in efficiency and effectiveness between the 2D and the 3D visualisation are minimal when compared generally (section 4.1.3.1) but also when testing different hypotheses regarding the appropriateness of data displays within virtual environments for the analysis of the data in relation to landform (section 4.1.3.2). The largest difference is in time in stage I where only the 3D visualisation was interactive and allowed navigation and thus required more time. On this basis, the hypothesis that participants may spend longer in 3D as they are thinking not only about the data but also about the landscape (Bleisch et al. 2008) and thus generate more complex hypotheses about the analysed data in relation to the landscape is rejected. In addition to this time difference, some differences between 2D and 3D are found in the insight wording. In stage IIa, the much higher use of A3 words (form words such as steep, lower and highest) in 3D compared to 2D may indicate that participants benefit from the 3D visualisation which is in accordance to their comments about enabling them to see the variation in topography better (section 3.6.4). On the other hand, the low use of A1 words (exact altitude values such as 1950m or 2400m) in 3D points to the problem of label readability in not north-oriented environments.

Even though some differences in wording can be found, the preferences of the participants do generally not seem to influence efficiency and effectiveness of task performance. Neither the direct comparison of results in stage IIa for participants preferring either the 2D or the 3D visualisation (section 3.6.4) shows an influence nor is an influence visible in the overall comparison even though in total a larger number of participants commented positively about the 3D visualisation compared to the number of positive comments about the 2D visualisation (figure 3.59). However, visualisation preferences may influence or be influenced by the participant's motivation as indicated by the participants who like the 3D visualisation better. They reported a larger number of insights and spent more time with the visualisation (table 3.7). Thus, the work may be more satisfying for these participants but satisfaction is not directly measured in this research.

4.1.4. Tasks

The tasks along the research stages are designed to be of increasing complexity (sections 2.2, 2.5.1.2, 2.5.2.2, 2.5.3.2 and 2.5.4.2). In stage I two elementary tasks are defined to test if bar comparison in a virtual environment is possible at all and how accurately it can be done. In stage IIa seven tasks are set allowing evaluation of the appropriateness of displays comprising several single bars for data analysis. The influence of the seven tasks in stage IIa on the results as detailed below, leads to the use of a single synoptic task in stage IIb trying to avoid the influence of different tasks. No tasks are defined in stage III as data experts are expected to fulfil and report their own real world tasks to answer their research questions. The functional view of tasks defined by Andrienko & Andrienko (2006) differentiates between elementary and synoptic tasks. Elementary tasks deal with individual elements of data and are thus potentially simpler while synoptic tasks deal with sets of elements and references and are thus potentially more complex (Andrienko & Andrienko 2006). Additionally, for the task definitions in stages IIa and IIb, this functional view of data and tasks is well suited to separate between tasks referring either to location, altitude or both location and altitude. This separation is needed as a basis for answering the hypotheses regarding the analysis of the data in relation to landform and distribution (section 4.1.3.2).

The comments regarding task difficulty (section 3.6.2 and figure 3.56) show that the tasks of stage IIa are difficult to answer. It is the case that some tasks are easier to answer with certain stage IIa data sets and difficult or with no clear answer in other data sets. The more open task of stage IIb, which left the participants more freedom in the exploration of the data set by requiring them to report insights based on one general question, led to less remarks about task difficulty. The seven tasks (t1-t7) of stage IIa are of varying complexity (section 2.5.2.2) and participants needed varying amounts of time to complete them (figures 4.1 and 4.2, similar variation of task completion times with insignificant differences between 2D and 3D (Bleisch, Dykes & Nebiker 2009*b*)). Additionally, analysing the recorded answers regarding their wording and their reference to either location, altitude or a combination of location and altitude (sections 2.6.2.4 and 2.6.2.3) shows the strong influence of question wording on the reference of the answers (section 3.2.3.2, figure 3.17). The same influence is visible in the more detailed evaluation of answer wordings (section 3.2.3.6, figure 3.27). Asking for location or altitude information results in the use of location or altitude words respectively. When asking for location and altitude information both categories of words are used but location words slightly more often. It is even possible to differentiate in the wording of the answer if a question asked for location information before altitude information or vice versa (Bleisch et al. 2010*b*). It seems natural that asking, for example, for information about altitude results in answers using altitude words. However, in combination with the information that some participants thought the tasks difficult to answer (section 3.6.2 and figure 3.56) this may also be interpreted that participants sought to answer the tasks as closely to the question as possible even using similar words. It happened that participants stated that there is no answer to a specific task in a specific data set (section 3.6.2) but generally participants tried to answer the tasks. Thus, the wording of the tasks may have an influence on the answers. Trying to avoid this potential influence, stage IIb was set up to include only one generic task asking to report insights into the data in relation to location, altitude and time of the day. The question tries to balance the wording to give each aspect (the references location and altitude and the variable time of the day) the same weight. This gives the participants more freedom to report anything they find and is not restricted to answer specific questions which may be difficult to answer with a specific data set.

Comparing the confidence ratings in stage IIa in light of the different task reference sets (location,

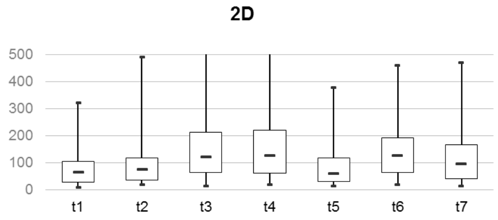


Figure 4.1.: Boxplots of task performance times per task (t1-t7) in seconds in 2D (max value t3=664, max value t4=1016)

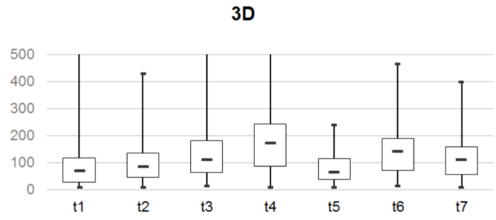


Figure 4.2.: Boxplots of task performance times per task (t1-t7) in seconds in 3D (max value t1=809, max value t3=830, max value t4=619)

altitude or both) shows a significant variation between the ratings for the three reference sets in 2D (section 3.2.3.3, figure 3.18). There are clearly less high confidence ratings of answers in the tasks referring to both location and altitude in 2D. Evaluating data in regard to location and altitude simultaneously results in less confidence in the answers than the other 2D tasks referring either to location or altitude. However, the variation between 2D and 3D for the tasks with combined location and altitude reference is not significant. This indicates that the more complex tasks with a combined reference set are less confidently solved in 2D and 3D.

The typical research questions reported by the data experts in the case studies (table 3.3) are analysed and typical tasks defined which would be performed in the visualisations to answer the questions. Comparing these 'real world' tasks to the tasks set in stages I, IIa and IIb show that they are similar. Typical 'real world' tasks include looking for particularly small or large values, comparing several values among each other to find patterns, relate single values or patterns of values to the landscape or comparing specific landscape features with the data occupying that landscape feature. Thus, the tasks performed in the experimental settings of stages I, IIa and IIb are representative of real world tasks even though they might not cover the whole range of possible tasks required to answer real world research questions.

Summary: The tasks along the research stages I, IIa, IIb are of increasing complexity. The real world tasks reported by the data experts in stage III comprise elementary and synoptic tasks and are thus of varied complexity. Comparing the reported real world tasks of stage III to the tasks performed in stages I, IIa and IIb shows that the latter are representative of the real world tasks even though they may not cover the whole range of tasks required to answer real world research questions. The separation of tasks referring either to location, altitude or a combination of both is enabled by Andrienko & Andrienko's (2006) functional view of data and tasks. It allows the hypotheses regarding the analysis of the data in relation to landform and distribution to be answered (section 4.1.3.2 and table 4.5) and is thus a valid and functional structure. The use of seven tasks in stage IIa compared to one task in stage IIb is discussed more detailed below (section 4.1.6). The evaluation of the collected data in stage IIa shows that different tasks with varying degrees of complexity influence the recorded answers. This effect is noted in differences in answering times, answer wording and references and also confidence rating between tasks. While these effects are strong between tasks in 2D and 3D, no difference can be found for the same tasks between 2D and 3D. This adds to the finding that the differences between 2D and 3D visualisations are minimal (section 4.1.3). The varying confidence ratings between tasks show that synoptic tasks referring to reference sets (location and altitude combined) are less confidently answered compared to elementary tasks referring to a single reference (location or altitude).

A number of other studies on 2D and/or 3D displays report findings which are to some degree dependent on the tasks. However, most such studies are domain specific. For example, Wickens et al. (1997) summarises several studies stating that the benefits of 3D displays are dependent on the exact spatially relevant aviation tasks. A more generic approach is taken by Tory et al. (2006) who conclude that precise tasks are most effectively performed in combined 2D/3D displays. Other study results may be task dependent but it is not obvious as they either comprise only a single task or were not evaluated for the influence of different tasks.

4.1.5. Settings and data sets

The variation of different efficiency and effectiveness measures between the 3D virtual environment and the 2D representation as discussed in section 4.1.3 is generally quite small and insignificant. Looking at the different settings and data sets we find stronger, for some aspects even significant variation. The influence of settings and data sets on efficiency and effectiveness of visualisations was not originally anticipated and thus not expressed in the original research aims and hypotheses (section 1.4). The evaluation of these influences is done mainly between stages IIa and IIb where basically the same data sets were used and thus direct comparison is possible. Additionally, as discussed in detail below, results of stage I and statements of stage III data experts add to the finding that setting and data set influence efficiency and effectiveness measures.

The two aspects, settings and data sets, are actually not separable as each geodata set relates to a specific environmental setting. However, to allow for comparison between stages IIa and IIb the summarising aspect setting (the combination of all data sets within a setting) is included in the evaluation. Stage IIa comprises eight data sets related to two different environmental settings W and S (section 2.5.2.3) while stage IIb consists of two data sets (aggregations of the stage IIa data sets) related to the same two environmental settings W and S (section 2.5.3.3). The following paragraphs compare the different appropriateness measures between the settings W and S in 2D and 3D.

Task completion times: In stage IIa the variation of task completion times is significant between the settings W and S in 3D (section 3.2.3.4, figure 3.20). Completing the tasks in 3D in setting S takes significantly more time than doing the same in setting W. The same but insignificant trend is visible for 2D. In stage IIb no significant variation nor trends between 2D, 3D, W and S are detected (section 3.3.3.2, figure 3.39).

Complexity and plausibility of insights: The variation of insight complexity between 2D, 3D, W and S is not significant in stage IIa (section 3.2.3.4, figure 3.22) nor in stage IIb (section 3.3.3.2, figure 3.41). The variation of insight plausibility between 2D and 3D is significant for setting S in stage IIa (section 3.2.3.4, figure 3.23). In the 3D visualisation of setting S more medium and less highly plausible insights are reported than in setting S in 2D. The same trend is visible if the plausibility of the insights in setting S is compared to setting W both in 3D. However, the latter variation is just not significant.

In stage IIb the setting S in the 3D visualisations has fewest highly plausible insights compared to all other settings (section 3.3.3.2, figure 3.42). The variation between setting W and S is significant in 3D and a similar insignificant variation is detected in 2D. In 2D and 3D the number of insights with low plausibility is higher in setting S than in setting W.

User con dence: The variation of the confidence ratings in stage IIa (section 3.2.3.4, figure 3.21) are significant between W and S in 2D but just not significant in 3D. In stage IIb (section 3.3.3.2, figure 3.40) the variation in confidence ratings between W and S in 2D and the variation between 2D and 3D in setting W is significant. In stage IIa setting S has fewer ratings of high confidence in 2D and 3D than setting W. The same is true for the setting S in 3D and the setting W in 2D in stage IIb.

Relation of data and landform: The variation of the insight references (location, altitude or both) are significant between the settings W and S in both 2D and 3D in stage IIa (section 3.2.3.4, figure 3.24). Setting S in 2D and 3D has fewer insights which refer to location and more insights referring to a combination of location and altitude. In stage IIb only the variation between the settings W and S in 3D is significant (section 3.3.3.2, figure 3.43). Setting W in 3D has fewest insights referring to altitude while setting S in 2D and 3D has more insights referring to both location and altitude.

Wording of insights: In stage IIa the variation in the wording of insights is significant between 2D, 3D, W and S (section 3.2.3.6, figure 3.26). Setting S in 2D shows a larger number of L3 words while setting W in 3D has fewest A3 words (section 2.6.2.4 for word categories). Setting S in 3D shows a different pattern having fewer A1 words and more A3 words. In stage IIb the variation in the wording of insights is significant between 2D, 3D and S (section 3.3.3.3, figure 3.44) except for the variation between 2D and 3D in setting W which is just insignificant. The word usage patterns for the two settings W and S in 2D and 3D are quite varied. Similar to stage IIa, setting S in 3D uses least L3 words.

Data sets: Evaluating the eight data sets of stage IIa in 2D and 3D (section 3.2.3.5) shows variation in the measures task performance time, confidence rating, complexity and plausibility of insights and insight references (figure 3.25). Applying an informal rating to the variations (section 3.2.3.5 and table 3.1 for details) it is concluded that data analysis in the 3D visualisation leads to better results in data set S1, while data analysis in the 2D representation is more appropriate for the data sets W1, W2, W3, S3, and S4. For the data sets W4 and S2 data analysis in a combination of both visualisation types may be best (table 3.2) as both visualisation types are ranked higher than the other in regard to some of the evaluated appropriateness measures. However, this assignment of the more suitable visualisation type to the eight data sets is exemplary. It is difficult to derive a set of characteristics from the eight data sets which helps to decide if a data set should be analysed in 2D or 3D (table characterising the data sets in appendix B.2). For example, as concluded above, data set S1 may best be analysed in the 3D visualisation. Looking at the characteristics of this data set (table B.2) it is the most complex data set covering the largest area, featuring the largest altitude difference and comprising most data values including the largest maximum value. Data set S3 is second most complex, however, it was better analysed in the 2D visualisation. Comparing the visual impression of the two data sets S1 and S3, it is found that S1 contains some larger, more visually prominent values than S3 where the data is distributed more evenly. However, other data sets, for example S2, also contain visually prominent values. Nevertheless, comparing the characteristics of all data sets (table B.2) and the visual impression of them, some tentative suggestions can be made (as discussed below in the summary).

Summary: Comparing settings W and S, some indications and trends are found that the two settings are different. Summarising, setting S tends to have more insights of high and less insights of low complexity compared to setting W in 2D and 3D. On the other hand, the insights in setting S are of less plausibility, with more insights of low plausibility, and less insights of high plausibility in 2D and 3D, particularly in stage IIb. Additionally, S tends to have lower confidence ratings of answers with more medium and less high confidence ratings. However, comparing settings W and S in 2D in stage IIb shows an inverse trend. Regarding the insight references, in stage IIb location references dominate. In stage IIa setting S has fewer location references than setting W. However, generally S has more combined location and altitude references in stages IIa and IIb. Comparing the wording of insights (section 2.6.2.4 for word categories), in setting S generally more A3 and less A1, A2 and L1 words are used (except for stage IIb which shows an inverse behaviour for A1, A3 and L1 in 2D and A2 in 3D). Additionally, setting S has more L2 words in 3D and less L2 words in 2D and more L3 words in 2D and less of them in 3D.

Looking at the two settings W and S and the corresponding data sets and, thus, the characteristics of the landscape and the area the data sets cover, some differences in land-cover, spread of data points and topography are found (table characterising the data sets in appendix B.2, screenshots of all data sets in appendix B.1). Setting W has less variation in altitude and most setting W data sets have fewer data points. Thus, they might be less complex to analyse and also yield less complex answers. Most of the results summarised above support this impression. On the other hand, the data sets of setting S are denser and topography and land-cover of the area are more varied. The correlation of land-cover and altitude in setting S (lower and wooded vs. higher up and above tree line) may also explain the increased use of references which combine location (e.g. land-cover) and altitude (e.g. landform). In relation to these findings, Brath (1997) proposes a number of metrics to measure effectiveness of static 3D views which include data density, complexity and others relating to the static views he evaluates. Further research in this area should aim to establish a set of characteristics which may allow the judgement of 3D visualisation effectiveness for specific data sets.

Another influence which cannot absolutely be ruled out is the fact that the data sets of setting W do not naturally belong to the area W (section 2.5.2.3). They were spatially transformed to create another data set in a different area for comparison. However, it was carefully tried not to create an unrealistic data set which does not show any relation to the landscape. But it may be possible that the not naturally existing relation between data and environment, a relation the tasks ask to evaluate, has some influence on the answers in setting W. However, the tasks are general data analysis and comparison tasks including minimal context. Additionally, the participants knowledge about deer behaviour (e.g. in relation to the landscape) is not that of a data expert. No comment noted that the data does not seem to match the landscape or that the deer seems to behave strangely compared to the landscape.

Comparing the characteristics of the eight different data sets in stage IIa (table B.2) and the visual impression of these data sets the following tentative suggestions for a conclusion in which visualisation a data set should most appropriately be analysed are made. It seems that data sets which are less dense and show data values which are somewhat clustered in groups are better analysed in the 2D visualisation (e.g. data sets W1, W2 and S4, section 3.2.3.5 and table 3.1). However, clustering is not visible in data set S3 which is also suggested for analysis in 2D. The 3D visualisation is appropriate for larger and denser data sets (e.g. S1) with some larger, visually prominent data values. This classification seems also to explain why settings W3, W4 and S2 gain from analysis in both visualisation types. They are not clustered but less dense (suitable for analysis in 2D as defined here) and contain some larger, visually prominent values (suitable for analysis in 3D as defined here).

From the case study data in stage III (table 2.3 and section 3.5.2) it is found that the data expert of case 'Brienz' is least inclined to use the 3D visualisation for data analysis. While this may be explained by the data experts knowledge of the data and his preferences for numeric data analysis (section 3.5.3), it could also be influenced by the fact that the data set is the smallest (65 points measured during 20 years) and covering the smallest area of all three case studies (table 2.3) thus supporting the hypothesis stated above that 3D visualisations are more suitable for larger, more complex data sets. Additionally, in case 'Literature Atlas' the data expert spent more time looking at and explaining the more dense central Switzerland data set (covering an area of complex topography) than the less dense northern Germany data set (covering flat area). This is especially interesting, as the central Switzerland data set is older and better known than the other one and asking the data expert to find new insights should thus guide her rather to the less known data set. In the case 'Deer SNP' the two data sets are looked at with about equal frequency. Roughly they can be judged as being about equally complex. One covers a larger topographically varied area displaying four variables on the data graphics while the other covers a smaller topographically varied area displaying 4-8 variables on the data graphics.

The finding from stage I that very small bars are difficult to compare (Bleisch et al. 2008, ➪ Appendix A) (in 2D and 3D) is supported by a data expert statement that short bars are not clearly visible in the virtual environment and that it is easier to look for tall bars in stage III (table 3.5). While no such comments can be found in stages IIa and IIb it may nevertheless be that a number of small bars is more difficult to compare than larger bars. A large number of small bars (e.g. showing value 1) is available in all data sets of setting S (S1-S4).

From this discussion it is concluded that data sets especially and settings more generally influence the analysis results strongly. This effect is found in all stages IIa, IIb and III which employ different data sets. Lloyd (2009) has reported that even using real data sets but 'uninteresting' categories may greatly influence the outcomes of a research project. It is likely that generally evaluations of visualisations are strongly influenced by the specific data sets used. This finding especially affects the results of evaluations which are based on single data sets (e.g. Koua et al. 2006) where without additional evidence the results may not be generalisable but only be valid for the specific data set used. Even a similar data set with the same variables may differ in the actual data values. The above discussion and categorisation of the eight different data sets of stage IIa (all data sets showing data from the same deer but at different times of the day) shows that while they all are thought similar they yield different results in terms of efficiency and effectiveness in the two visualisation types tested. Other evaluations employ known 'test data sets' for their evaluation of displays (e.g. the cars data set, Pillat, Valiati & Freitas 2005) but even then the influence of this specific data set may not be negligible when the different visualisations are only compared on the basis of a single data set.

4.1.6. Increasing complexity - comparing stage IIa and IIb

The research objective of stage II (stages IIa and IIb) asks for an evaluation of the appropriateness of visually combining quantitative data with virtual environments for more complex visualisation tasks. While the results on this have been discussed in combination with stage I and III results in section 4.1.3 above, some more details can be found by comparing stages IIa and IIb directly. As described in the methods chapter (section 2.1.0.4) stage II is subdivided in the two stages IIa and IIb. In stage IIa, task complexity is increased by employing elementary and synoptic tasks while still evaluating single bars (univariate data) similar to stage I. In stage IIb data complexity is increased by introducing bar charts (multivariate data, as normally available in real world settings such as tested in stage III) but basically employing the same data set as in stage IIa. A comparison of the stage IIa and IIb results allows a direct observation of effects caused by increased data complexity. Additionally, while both stages are methodologically similar, slight changes in method were introduced from stage IIa to stage IIb based on the analysis of the stage IIa data (e.g. single exploratory task and insight reporting in stage IIb compared to seven tasks and answer recording (split to insight during analysis, figure 4.3) in stage IIa, sections 2.5.2 and 2.5.3). Thus, the effects of these method modifications can also be analysed when comparing stages IIa and IIb. While these two types of effects are not completely separable during analysis, it is possible to follow a replication strategy (also including information from the other stages) and finding the most suitable explanation amongst rival explanations (Miles & Huberman 1994).

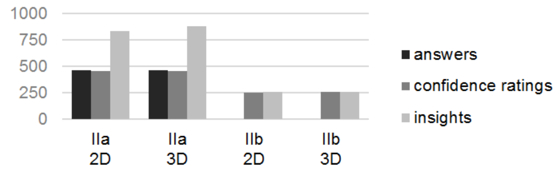


Figure 4.3.: Number of recorded answers (split to insights based on their content, section 2.6.2.1) and insights in IIa and IIb, confidence ratings are recorded per answer (IIa) or insight (IIb)

User con dence: User confidence is recorded per answer in stage IIa and per insight in stage IIb (figure 4.3). Thus comparison is pro-rata for mean number of insights per answer in stage IIa and indicative only. The variation in the confidence ratings between stage IIa and IIb in 2D and 3D is not significant (section 3.4.0.2, figure 3.46). The confidence ratings show that generally more ratings of high confidence are recorded in 3D (in stage IIa and IIb). Additionally, participants in stage IIb are more confident in the reported insights (more ratings of high confidence and less ratings of low confidence) than participants in stage IIa. This may reflect the fact that participants in stage IIb are unguided by questions they found difficult to answer (section 3.6.2 and figure 3.56) thus they may only report the insights they have more confidence in.

Task completion time: The mean time needed to find and report an insight is compared roughly between the stages IIa and IIb (section 3.4.0.1, figure 3.45). Each answer of stage IIa contained on average 1.79 (in 2D) or 1.91 (in 3D) insights (figure 4.3). The mean time spent on finding and reporting an insight is longer in stage IIb than in IIa (1.6 times the time for 2D and 1.7 times the time for 3D).

Insight complexity: The variation in the complexity of insights between stage IIa and IIb is significant in 2D and 3D (section 3.4.0.3, figure 3.47). The data displayed in figure 3.47 shows that there are more medium and highly complex insights but less insights of low complexity reported in stage IIb compared to stage IIa. These results reflect the increase in complexity that is expected from the design of stages IIa and IIb. Additionally, these results are to some degree dependent on the definition of complexity as used in this study (section 2.6.2.1 and table 2.4) where the analysis of more complex data with more variables as used in IIb leads, potentially, to more complex insights. But this definition of complexity is accepted as valid as an insight which, for example, comparatively analyses several attribute values is considered as being more complex than an 'insight' which describes the variation of a single attribute. Thus, according to this definition the participants in stage IIb have, potentially, a better chance to report more complex insights as the data analysed in stage IIb is more complex. However, even there simple descriptions of low complexity are possible if a participant is not able to compare the data and find more complex insights. Additionally, insights are rated to be more complex if they include context information or conclude in some sort of hypotheses (section 2.6.2.1 and table 2.4). In stage IIb more insights which fit in this category are reported.

Insight plausibility: The variation in the plausibility of insights between stage IIa and IIb in 2D and 3D is not significant (section 3.4.0.4, figure 3.48). Figure 3.48 shows that the plausibility ratings are very similar in 2D and 3D and in both stages IIa and IIb with slightly less insights of high plausibility in stage IIb.

Relation of data and landform: The variation in quantity of insights referring either to location (L), altitude (A) or both (LA) between the stages IIa and IIb is significant in 2D and 3D (section 3.4.0.5, figure 3.49). Location (L) is clearly more often used as an insight reference in IIb than in IIa.

Wording of insights: The variation in the quantity of words in the categories L1-L3 and A1-A3 (section 2.6.2.4 for word categories) between the stages IIa and IIb is not significant in 2D but is significant in 3D (section 3.4.0.6, figure 3.50). In stage IIb 2D insights used more words in general and, similarly to stage IIb, 3D insights used more A1 and A2 but clearly less A3 words than 2D and 3D insights in stage IIa. Especially, 3D insights in stage IIb used fewer L1 words and 2D and 3D insights of stage IIb used fewer L2 but more L3 words than insights in stage IIa. This is connected to the finding reported above that in stage IIb participants reported insights with location reference more often.

Summary: The main difference between the two stages are the multivariate data displays and single synoptic question in stage IIb as compared to the seven elementary and synoptic questions and univariate data displays in stage IIa. While the data and task complexity is increased between stage IIa and IIb a modification in methods is also introduced in stage IIb where a single task and insight reporting is employed compared to answering seven tasks in stage IIa. The influences of increased complexity and method modifications on the results as compared above cannot always clearly be separated.

The results reported above indicate that participants in stage IIb, unguided by questions they found difficult to answer (section 4.1.4), reported insights with higher confidence and potentially under less time pressure as they take more time to find insights in stage IIb and thus report less of them in the same time. As stage IIb the aggregated data sets of stage IIa shows there is the potential of finding at least the same number of insights if not more. Even though the questions of stage IIa are thought difficult, they provide some guidance and result in more reported insights as participants tried to answer the questions even though some questions may be more difficult to answer in some of the data sets (section 4.1.4). Looking at the plausibility of the insights in stage IIa they do not seem to be of lower plausibility. It rather seems that participants of stage IIb reported only the more obvious results of the data analysis where they have some confidence in or they got bored more quickly in analysing the data sets without specific tasks to answer. In some cases more trivial insights reported towards the end of the data analysis time (about 1/2 hour) can be found when evaluating the participants reports of insights indicating that indeed some participants seemed to get bored and do less comprehensive analysis of the data the longer they work with it.

Additionally, the aspect of data graphics complexity may influence the number of insights recorded. In stage IIb each data graphic showed up to eight different variables (typically 1-4). As discussed above (section 4.1.1) participants prefer data graphics showing not more than 3-4 variables. Thus, the more complex, or too complex, data displays in stage IIb may have hindered detailed data analysis and the reporting of more insights. However, the largest part of the data graphics in stage IIb showed 1-4 variables and were thus below the identified complexity threshold.

Looking at the complexity and plausibility of the recorded insights we find that insights in stage IIb are generally of higher complexity which reflects the designed increase in data and task complexity between the two stages IIa and IIb. Additionally, this is also dependent on the definition of complexity used in this study. However, analysing the content of the insights of high complexity shows that participants of stage IIb tried more often to include context information such as their own (potentially limited as they are not data experts) knowledge of deer behaviour. Insights containing statements like "on the green bit for browsing" or "on the eastern slope to warm up in the morning sun" are reported. Thus, the task of reporting insights in regard to one general question seemed to leave the participants more freedom to think generally about deer behaviour as compared to stage IIa participants who seem busy to answer the tasks. In contrast to the more complex insights in stage IIb the insights are of slightly less plausibility (insignificant variation; less highly plausible insights but about the same number of insights of medium plausibility and in 3D an even lower number of insights with low plausibility). To some degree this might be explained by the above statements. Participants trying to include context knowledge even though they are not data experts may be musing about less plausible deer behaviour. While the researchers analysing the data were not data experts either they have a better knowledge of the data as they were working with all data sets and for a much longer time. Thus, they have potentially rated some insights including deer behaviour musings as less plausible.

The most striking difference between stage IIa and IIb can be found in the wording of insights and the insight references. While in stage IIa in 2D and 3D all three types of insight references (location, altitude and combination of location and altitude) are used about similarly often, in stage IIb insights referring to location are about three times as frequent as insights referring either to altitude or location and altitude combined (in 2D and 3D). This shows that the questions of stage IIa definitively guide the answers. The general question of stage IIb included location and altitude reference to the same weight but the insights reported reflect the freedom of the participants. It suggests that participants in general are not used to analysing data in relation to landform. They do so if asked to do so as is the case in stage IIa where the questions explicitly asked for data and landform comparison. But they report at the same time (in stage IIa) that the questions are difficult to answer (section 4.1.4). If they can generally report their insights they revert to reporting mainly insights relating to location and distribution which they might be more used to doing even though the question also asked for insights in relation to altitude. Insights referring to altitude or a combination of location and altitude are about half as frequent in stage IIb as in stage IIa. A statement from the data expert in the case 'Deer SNP' supports this conclusion. She reported that their main data analysis is in regard to location and distribution even though altitude and landform are interesting. Analysis regarding the latter is done only in special cases and it is more complicated to do.

4.1.7. Further visualisation aspects

Suitability: The participants comments on suitability (section 3.6.4) show that in the more complex stage IIb 3D is more often commented upon as being better than 2D compared to the less complex stage IIa. This may add to the suggestion (section 4.1.5) that 3D is better for more complex data sets and settings. However, this finding is not free of the confounding factor 'more freedom' in insight reporting in stage IIb based on one task compared to answering seven tasks in stage IIa (section 4.1.6). This may lead to more positive comments about the 3D visualisation in stage IIb. The suitability of the visualisation is judged quite positively in stage III (section 3.5.4).

Some of the participants who prefer the 2D visualisation also commented in defence of their map reading skills (section 3.6.4). Another study (Bleisch & Dykes 2008) has found that self-reported map reading skills are generally high and that participants tend to defend these. The group of informed users who participated in this research (section 2.4) is thought to have at least basic but probably better map reading skills. In stage I, the testing of the participants' general spatial ability with the

Santa Barbara Sense of Direction test (Hegarty et al. 2002) allowed confirmation that the informed users generally have high spatial abilities (Bleisch et al. 2008, ➪ Appendix A).

Previous visualisation usage: In stage III the type of previous visualisation usage influences greatly how the data experts would like to use the presented representation of quantitative data graphics within virtual environments (table 4.6). Data experts who have used visualisations mainly for communication purposes or for illustrations would like to use the presented 3D visualisation the same way, while the data expert who already used visualisations mainly for data analysis purposes would like to do so with the presented 3D visualisation as well. Some experiences and comments from stages IIa, IIb and III of this research (e.g. comments asking for other visualisations which 'show' the data variation better (e.g. density maps), are easier to interpret or are sort of pre-interpreted (e.g. curvature maps), section 3.6.3, or comments stating that the visualisations can only be used to get an overview while detailed analysis needs to be done elsewhere) together with the finding that people find it difficult to relate data and landscape (as reported for answering the stage IIa tasks) leads to the general question of how often data visualisations are really used for exploratory data analysis and not only for communication purposes. MacEachren (1994) presented the predecessor of the geovisualization cube (which was re-used and slightly re-designed over time, e.g. in MacEachren & Kraak 1997) which provides a framework for categorising different visualisations according to intended use (public or private), level of interactivity and the type of data to be shown. These visualisation categories range from visualisation or exploration of unknown data to communicating known facts. Newer, large scale initiatives (e.g. 'visual analytics' in the USA (Thomas & Cook 2005) or 'VisMaster' in Europe (Keim, Kohlhammer, Ellis & Mansmann 2010)) promote exploratory data analysis with visualisation. From the findings in this research it seems necessary to promote the use of visualisations for data exploration and analysis. Visualisation users (not visualisation researchers) in their 'daily business' seem to use data visualisations rather for communication and illustration purposes than for more complex data exploration and analysis tasks.

		stage III (section 3.5.3)
exploratory question	✔	Q III.III: Can differences in appropriateness of the visualisation for data analysis be related back to different characteristics of the evaluated application settings?

Table 4.6.: Overview of explored question (✔ = accepted, ✘ = rejected)

Visualisation completeness: Another aspect that was found while exploring the question QIII.III in stage III (table 4.6) is the issue of showing all essential data in the same display. Not showing all important data aspects (e.g. the important bearings of the slope movement vectors in the case 'Brienz') reduces the perceived usefulness of the visualisation. From the three case studies it seems that the more important the not displayed data is the less useful the visualisation becomes. This additionally shows that the more aspects are controlled for in a visualisation evaluation (e.g. in experimental settings) the less economically valid the results may be.

4.2. Re ection on the methodological framework

The research designs and employs a sequential mixed methods approach with four research phases (stages I, IIa, IIb and III) which bridge between 'in vitro' and 'in vivo' research approaches. Additionally, quantitative and qualitative research methods are also used in combination within most of the research stages. Thus, an across-stage and within-stage mixed model research design is employed

(Johnson & Onwuegbuzie 2004). Details about the employed methods and definitions in general and for each research stage separately are available in chapter 2 (table 2.1 for an overview).

The following sections reflect upon the methodological approach of this research. A short summary of each research stage is provided before reflecting upon selected methodical aspects and the methods and methodological framework in general.

4.2.1. Summary of the research stages

Stage I:
The goal of stage I is to test whether the data graphic type 2D bars on billboards as derived from literature about 2D representation methods are appropriate for the display of quantitative data in virtual environments. An experimental setting is employed in stage I to test the efficiency and effectiveness of the data graphics for two elementary tasks by judging pairs of bars representing two different values. The tasks are identifying the taller bar and comparing the length of two bars. In a 2x2 factorial within-subject design 26 participants judged a subset of 20 different bar combinations in four different settings (2D with and without frames, 3D with and without frames). They participated in a laboratory face-to-face setting completing a digital questionnaire. The quantitative analysis of task performance time, identification of the taller bar and bar length differences allows the proposed hypotheses to be accepted or rejected (table 4.7). It is concluded that difference between 2D and 3D lies rather in efficiency than effectiveness. In the 3D visualisation the reference frames of the bars improve task performance times.

		stage I (Bleisch et al. 2008, ➪ Appendix A)
hypotheses	✔	H I.I: Users are able to identify the taller of two bars in a 3D desktop virtual environment as well as they can in static 2D graphics.
	✔	H I.II: The effectiveness of estimating differences between two bars is not significantly different in the 2D and 3D settings (2D without frame, 3D with frame).
	✘ ✔	H I.III: The efficiency of task completion is improved by the use of a reference grid (frame) (in the 2D ✘ and 3D ✔ settings).
	✘	H I.IV: The efficiency of estimating absolute values from bar lengths (with a reference frame) is not significantly different in the 2D and 3D settings. (3D takes longer)

Table 4.7.: Overview of tested hypotheses in Stage I (✔ = accepted, ✘ = rejected)

Stage IIa:
The goal of research phase II is to test whether 2D bar displays in virtual environments are appropriate for more complex visualisation tasks involving landform. Stage IIa employs an experimental setting to test more dense univariate data sets with more complex tasks. Seven tasks are defined which include two elementary tasks and five synoptic tasks all referring either to location, altitude or a combination of location and altitude. A balanced within-subject design is employed to assign four of the eight data sets in either the winter or the summer setting to the 34 participants. The participants took part via an online questionnaire containing explanations, training settings and the actual tasks to be solved in the different data sets. The quantitative and qualitative analysis of task performance times, task answers, confidence ratings and comments allows rejecting the proposed hypotheses (table 4.8). It is concluded that, in terms of efficiency and effectiveness, neither the 3D visualisation does improve the relation of data to landscape, nor does the 2D visualisation improve the relation of data to location and land-cover. Both visualisation types perform similarly well. However, efficiency and effectiveness are influenced by data set, setting and task (section 4.1.4 and 4.1.5).

		stage IIa (section 3.2.2)
hypotheses	✘	H IIa.Ia: Tasks where the reference set is location are solved more efficiently in the 2D representation.
	✘	H IIa.Ib: Tasks where the reference set is location are solved more effectively in the 2D representation.
	✘	H IIa.IIa: Tasks where the reference set is altitude are solved more efficiently in the 3D virtual environment.
	✘	H IIa.IIb: Tasks where the reference set is altitude are solved more effectively in the 3D virtual environment.
	✘	H IIa.IIIa: Tasks that include combined reference sets (location and altitude) are solved more efficiently in the 3D virtual environment.
	✘	H IIa.IIIb: Tasks that include combined reference sets (location and altitude) are solved more effectively in the 3D virtual environment.

Table 4.8.: Overview of tested hypotheses in stage IIa (✔ = accepted, ✘ = rejected)

Stage IIb:
The goal of research phase II is to test whether 2D bar displays in virtual environments are appropriate for more complex visualisation tasks involving landform. Stage IIb employs an experimental setting to test multivariate data sets with more complex tasks. One synoptic task is defined which refers to location, altitude and attribute space in a balanced way. A balanced within-subject design is employed to assign the two data sets in either the winter or the summer setting to the 38 participants. The participants took part in a laboratory face-to-face setting. They first trained in the interaction with the display and then reported insights regarding the task in a digital questionnaire. The quantitative and qualitative analysis of insight reporting times, insights, confidence ratings and comments allows the proposed hypotheses to be accepted or rejected (table 4.9). It is concluded that the 2D and the 3D visualisations are similarly efficient and effective also with more complex data sets. However, performance is influenced by data set and setting (section 4.1.5).

		stage IIb (section 3.3.2)
hypothesis	✘	H IIb.Ia: Insights into the relationship between data and location/land-cover/etc. are more often and more efficiently reported in the 2D representation.
	✘	H IIb.Ib: Insights into the relationship between data and altitude/landform are more often and more efficiently reported in the 3D virtual environment.
	✔	H IIb.II: The complexity and plausibility of insights reported do not vary significantly between the 2D representation and the 3D virtual environment.

Table 4.9.: Overview of tested hypothesis in stage IIb (✔ = accepted, ✘ = rejected)

Stage III:
The goal of stage III is to explore whether 2D bar displays in virtual environments are appropriate in applied scenarios with real tasks and data experts. The propositions and research question are explored by employing a multiple case design and purposively selecting diverse cases. Such settings include the whole wealth of context information which was left out or controlled for in the other stages. In exchange, controlling for context allowed for the participation of a larger group of users. In stage III, the three data experts provided data sets, descriptions of the data set and context information in a digital questionnaire. Different 3D visualisations were created and discussed with the data experts at their workplace according to the case study protocol. The collected data is qualitatively analysed and reported as 'thick matrix displays' (section 2.6.2.5) and summarising evaluations in regard to the research propositions and questions (table 4.10). The design of the case studies did not anticipate that the data experts are more interested in the visualisation possibilities for future analysis than in (re-)analysing the data sets they know (very) well (section 3.5.2). Thus, the case studies as done in this research, with known data sets and rather short-term researcher data expert relationships, do not allow the determination of whether data experts gain insights into their data sets. For the same

reasons, the data experts did not fulfil the sort of real world tasks anticipated and reported by the typical research questions of each case study. But for comparison purposes, the reported typical research questions translate in the type of tasks set in the previous research stages. Despite the shortcomings, the results from the case studies give valuable insight into various aspects of the 3D visualisations and in the appropriateness of them for real world settings. This is strengthened by the fact that this is the only research stage which, intentionally, includes the whole wealth of context information. From the data analysis it is concluded that the importance of relation to the landscape does not have an obvious influence on the appropriateness of the visualisation type. More influential are the completeness of the display and the data expert's previous usage of other visualisations.

		stage III (section 3.5.3)
propositions	✘	P III.I: Displays of quantitative data within virtual environments are more appropriate for the visual analysis of the data if the data and landscape (2D/3D) relationship is more important in an application.
	✳	P III.II: A visual combination of data and landscape allows more efficient and/or more effective finding of insight into the data.
	✔	Q III.III: Can differences in appropriateness of the visualisation for data analysis be related back to different characteristics of the evaluated application settings?

Table 4.10.: Overview of evaluated propositions and questions in stage III (✔ = accepted, ✘ = rejected, ✳ = not answered)

4.2.2. Employing a mixed methods approach

The main reason for designing mixed methods approaches for research studies is the utilisation of "complementary strengths and nonoverlapping weaknesses" (Johnson & Onwuegbuzie 2004, p. 18). In an ideal case, the weaknesses of one method are compensated for with strengths from other methods while ideally even having strengths in similar areas thus increasing the validity of the research. In this research, a sequential mixed methods approach employing four research stages is designed. It is hypothesised that the main differences between 'in vitro' and 'in vivo' research are the differences in data and task complexity, number of relevant users and context (tacit knowledge of users and also the number of controlled for or allowed influencing factors) as illustrated in figure 2.1. Thus, the four research stages explore the issue of quantitative data displays in virtual environments by bridging from 'in vitro' controlled experimental research methods to 'in vivo' context-rich case studies in real world applied settings. Each of the identified research stages uses its own research method(s) and the results are used to inform the following stage(s). While the collected data of each stage is mainly analysed separately, it is evaluated integratively across all stages for certain aspects. Additionally, the results and findings of all stages are integrated in the discussion. This allows conclusions to be drawn forward and backward along the research stages. While in the forward process findings from an earlier stage inform later research stages backward inferences are done by revisiting and re-evaluating earlier collected data and findings in the light of later findings. The main purposes for conducting mixed methods research in this study is complementarity (to elaborate and enrich findings from one method with results from another), expansion (aiming for breadth) and triangulation (seeking convergence of findings) (Greene et al. 1989) in accordance with the research aim of increasing understanding of the use of desktop-based virtual environments with a focus on the graphical representation of quantitative data through abstract data graphics.

Before reflecting generally upon the mixed methods approach and methodologically bridging between 'in vitro' and 'in vivo' research methods a number of research aspects are methodologically evaluated. It is described how the combination of methods along the bridge complement, expand or triangulate the findings in regard to a specific research aspect.

Data graphics:
Based on literature about 2D representation methods the 2D bars on billboards with and without reference frames are hypothesised to be appropriate for the display of quantitative data in virtual environments (section 2.3.1). In stage I, this hypothesis is tested in an experimental setting with elementary tasks and no context. Positive results of elementary tasks employing a somewhat psychophysical approach by measuring responses to a changing stimulus (bar length) are required as basis for more complex evaluations (Montello 2002) even though they are only limitedly generalisable to more complex data and tasks. The results of stage I serve as input to all further research stages. Based on that results it is decided to use bars with frames in the 3D virtual environment and bars without frames in the 2D representations. In the course of employing a series of different research stages it is possible to make such a decision based on results from simple data and tasks even though they might not hold true for more complex data and tasks and, thus, do not completely reflect the appropriateness of 2D data graphic displays within virtual environments. Evaluating the data from all stages it is then found that the use of bars or bar charts with frames can be crowding and confusing in more data dense displays in 3D. Additionally, a complexity threshold of maximally 3-4 variables per data graphic is identified which participants do not like to see crossed (section 4.1.1). The finding that participants principally are able to compare bars in virtual environments held true through all research stages. But findings from controlled context-free experiments as employed in stage I need further evaluation and interpretation before they can be applied to real world settings. North (in Jankun-Kelly 2006) remarks that it is difficult to model high-level visualisation purposes with results of low-level task performance measures. In this research, the overall findings in regard to the data graphics are stronger and more detailed because of the re-evaluation of the stage I low-level task performance results in all of the following increasingly applied stages. The re-evaluation is useful as limitations, such as the crowding reference frames and not liked complex bar charts, do not show up until stage IIb and III.

Participants:
The participants of stages I, IIa and IIb are current and former students and staff members of geography and geomatics departments at three different universities in Switzerland and the UK. They are thought to constitute an informed group of typical users with some knowledge about data visualisation and some 'geo' background. Potentially, each of them could become a data expert in some knowledge area. The data experts from the case studies, on the other hand, also belong to the group of informed users with some data visualisation knowledge and 'geo' background or experience. This is tested by baselining the data experts with the stage I experiment (section 3.5.1). In stage I the participants were tested for their general spatial abilities with the Santa Barbara Sense of Direction test (Hegarty et al. 2002). This allowed confirmation that the informed users participating in stage I generally have high spatial abilities, as was aimed for (section 3.1.3). While the results of all the research stages are primarily valid for the tested group of participants the results from all research stages indicate that they may be applicable to a wider group of people with similar characteristics. The sequential design of the research stages where each stage is designed to at least partially test the findings of the previous stage helps improving the validity of the selected group of participants.

Marsh (2007) concludes that students should not be used as substitutes for visualisation or domain experts. In this research, the last year geomatics students have similar characteristics (e.g. education in information visualisation and geographic information) as many of the potential data experts, even though, they do not have the context knowledge of a specific application area. The evaluation of the data from all stages, but especially from stage IIb and III, shows that the students or rather informed users participating in stage IIb commented upon many issues which are later reinforced by statements of the data experts in the case studies (e.g. data graphic complexity). Informed participants, as took part in this research, are not data expert substitutes. But they may help to detect issues in visualisation evaluations as they are more available than data experts. Data experts may be called upon for validation and further exploration. Some of the information contained in the stage III data set (e.g. the importance of visualisation completeness or the influence of high familiarity with a data set) could not be gathered in earlier research stages (section 3.5.2). In stages IIa and IIb, participants tried to include their (potentially) limited knowledge about deer behaviour. This led to insights which were rated with higher complexity (including more context information, section 2.6.2.1) but also potentially rated with lower plausibility because of unrealistic deer behaviour musings. In future evaluations using data experts and informed participants in stages close to real world settings (e.g. stage IIb) should be considered. Comparison of the results of the two groups of participants may yield more differentiated results on the influence of context information.

While all participants in this research are thought to build a group of informed users none of them is 'average'. Their individual differences, which also exist for a group of users characterised as being similar, are not detailed in the evaluations. However, these differences are visible to some degree in the collected data, for example, in the range of times needed to answer a task or report an insight, in the differences in confidence ratings and in the variations in complexity and plausibility of answers. Some of them state that they prefer either type of visualisation (section 4.1.3) and two of the data experts report that they are not too confident into their bar comparison results (stage I tasks, section 3.5.1). For all participants together no significant and relevant differences in the appropriateness measures between 2D and 3D visualisations are detected. However, there may be individual differences which are not looked at. Thus, as no single participant of the group of informed users may be exactly 'average' (cf. Yau 2011, for the 'average' or most typical person in a more general context) the findings may only to some degree be valid for each single participant. However, for the group of tested participants they are valid and their validity is even strengthened by testing a larger group of participants with similar characteristics in different experimental and applied settings.

Tasks:
The use of tasks in the different research stages is discussed in detail in section 4.1.4. In stage IIa an influence of task definition on the results was found. It was thus decided to use only one synoptic task in stage IIb. This change together with changing from answering tasks online to reporting insights in a face-to-face setting between IIa and IIb resulted in a number of interesting results, such as participants taking more time to report single insights, which are discussed in section 4.1.6. The influence of task definitions in stage IIa on the results was not originally anticipated and thus also no hypotheses formulated to test for it. Answers influenced by tasks are not basically bad. It is expected that more complex tasks may take more time and that answers are guided by the tasks. However, the influence or guidance should not be negative in the sense that, for example, participants try answer where no answer exists in the data. The sequential mixed methods design allows accommodation of a design change from stage IIa (seven specific tasks) to IIb (one generic task). One flaw in the design change is the problem that differences in the results between stages IIa and IIb which should reflect increasing complexity are not completely separable from influences of the design change (section 4.1.6). Research methods, except the most controlled experiments, are often subject to confounding issues or influencing factors. Mixed methods approaches are designed

to mitigate such influences by comparing findings across different research methods and in this research especially across all research stages. For the task issue the advantage of the sequential mixed methods design is the possibility to use seven/one task in the two research stages. The research questions ask for a comparison of the performance between the 2D and 3D visualisations which can be done independent of the task type as long as the tasks are the same in the 2D and the 3D representation. Discussion of performance between 2D and 3D along the research stages is also independent of task type. However, a design change does additionally allow comparison of results for different tasks set which would not be possible by using a single method or research stage. As said, this is not free of confounding factors as also the data complexity is increased between stage IIa and IIb. To overcome these influences a replication strategy is applied and the most suitable explanation amongst rival explanations is looked for (Miles & Huberman 1994). Thus, for each difference in the result which is detected, not only the data from stages IIa and IIb is consulted but also the data from stages I and III. This showed that the reports about difficult to answer tasks in stage IIa are mainly a problem of the participants who are not used to analysing data in relation to landform and altitude rather than difficult to understand tasks, as was suspected first. In comparing stages IIa and IIb, it was also found that specific tasks do guide the participants but this can also have advantages, such as participants doing data analysis they do not do habitually (section 5.3.2). On the other hand, the result that in stage IIb more complex insights were reported than in stage IIa is best explained by the increased complexity of the data and the task and not by the change in design from answering tasks to reporting insights.

Data sets:
The data sets used for evaluation in the different research stages are of increasing complexity. Stage I employed 20 pairs of single values (random but controlled distribution). In stage IIa eight also univariate but real data sets are used. The same data is aggregated in stage IIb and displayed as multivariate data set. The real world data sets of the three case studies in stage III are multivariate. Testing first with single random values is a valid method. The results of stage I are indirectly confirmed by the results of the later stages. The same approach of testing first with elementary tasks has been used in another experimental study in the meantime (Bleisch accepted). In stage IIa and IIb the general comparison of the results between 2D and 3D shows no significant differences. However, data sets and settings (the area a geographic data set covers) have strong influences (section 4.1.5). In stage IIa it is shown that some data sets may be better analysed in 3D visualisations others in 2D visualisations. Based on the data set characteristics in stage IIa it is tried to suggest which data sets should better be analysed in 3D and which in 2D. The consultation of stage III information on data sets and characteristics strengthens the suggestion.

The increasingly complex data sets are also reflected in the increasingly complex data graphics. In stage I and IIa single bars suffice to display univariate data while in stage IIb and III the multivariate data sets call for bar chart displays showing up to eight variables at once. The case study data sets in stage III are of varying complexity and also sometimes thematically aggregated (section 2.3 for examples of data graphics). Thus, the data experts were able to compare between simpler and more complex data graphics. They remarked that no more than 3-4 variables should be displayed at once. Based on this input the comments of stage IIb where the data graphics show up to eight variables are re-evaluated. The issue of too complex data graphics shows up in a few comments in stage IIb but would not have been considered sufficiently without the data experts input.

The selection of diverse cases which vary in different dimensions of interest in stage III allows covering a wide field of potential real world applied settings and data sets, even though they may not be representative of all real world cases. Additionally, as the results show, the data collected in stage III helped interpretation and validation of issues arising in earlier stages while simultaneously allowing the breadth of the research to be expanded through exploration of further issues such as visualisation completeness or further usage of quantitative data visualisations within virtual environments. The data set of the case 'Brienz' may be too small to need visual analysis of the data. However, slope movement observations are common data sets with a close relation to landform even though other monitored areas may be even smaller.

Context:
The experiment in stage I tried to control as many influences as possible by defining the settings and tasks exactly. It is context-free as neither the evaluated data nor the background landscape has any meaning. In stage IIa and IIb the participants are given some information about the data and setting. They have tried to include their (potentially limited) knowledge of deer behaviour in the task answers and reported insights. Stage III is the only research stage which, intentionally, includes the whole wealth of context information and influencing factors and does not try to control for, but rather tries to capture, information about the context. Thus, the design is not based on tasks which need to be completed such as the earlier stages but rather on capturing the information while talking with the data expert about the data sets, the context and the 3D visualisation. While this potentially allows information to be captured in breadth and depth, it is difficult and also not acceptable to guide the data expert. Data experts not re-analysing their data, thus, cannot and do not want to be forced to do so. The issues of visualisation completeness (section 3.5.3) which arose during the analysis of the stage III data emphasises that the more controlled a setting is (e.g. the stage I experiment with selected data and no context) the less economically valid it potentially is. Thus, the combination of research stages is valuable to gain insight into research aspects evaluated experimentally 'in vitro' and via case studies 'in vivo'.

4.2.3. Further methodological aspects

Experimental balanced within-subject designs:
The experimental methods of stage I, IIa and IIb (sections 2.5.1, 2.5.2 and 2.5.3) implement balanced within-subject designs (Elmes et al. 2006). Thus, the different settings, dimensions and tasks are assigned to participants in various predefined orders to minimise confounding carry-over effects such as familiarisation with a setting, getting used to the tasks or getting tired towards the end of doing several tasks. For stage IIa it is evaluated if the order of the different displays used really has an effect and thus the balancing within-subject design of the experiment is sensible. The analysis shows that there are significant differences of the confidence ratings for the first or the second display of the same visualisation type (section 3.2.3.3, figure 3.19). The second settings of both 2D and 3D representation show more high confidence ratings. Additionally, while the second 2D representation shows less low confidence ratings the second 3D visualisation shows more low confidence ratings. Thus, some participants seem to get more confident during the course of the experiment while others may get less confident while answering the tasks using several visualisations. It makes sense to use a balanced design which averages such effects over all data displays tested.

Measuring appropriateness:
The appropriateness of quantitative data displays in virtual environments is evaluated through measuring efficiency and effectiveness of the visualisations. While efficiency is analysed through task completion time, measures of effectiveness are evaluated based on the concepts of accuracy and

utility (section 2.1.0.2). In stage I it is possible to measure the accuracy of task completion. In stage IIa and IIb efficiency is measured through the analysis of the reported insights in regard to complexity, plausibility and user confidence (Rester et al. 2007). The results show that these measures are basically valid as they allowed the proposed hypotheses to be accepted or rejected. However, some issues in regard to insight evaluation arose which require discussion.

In implementing stage IIa the participants were expected to answer each of the seven tasks with a short statement which can be regarded as one insight (based on the definition of insight by Saraiya et al. 2005). However, participants took about 1.4 times the time than was anticipated to complete the questionnaire and they gave very detailed answers which cannot be regarded as a single insight but rather as an insight report. Thus, these reports needed splitting to single insights which could be rated as intended. As the time is recorded per answer and there is no way to split it up per insight (e.g. splitting up pro rata of the number of characters of each insight does not take into account that the process of thinking about an insight may take much longer than what is reflected in the number of words or characters used to report it) it is only possible to directly compare task completion time within a research stage (e.g. between the 2D and the 3D visualisation of stage IIa). Comparison between research stages is only roughly possible. In stage IIb, were participants were also asked to report single insights, some lengthier insight reports needed splitting. While in stage IIb this is more rarely the case than in stage IIa, the same issue regarding time applies. Identical to the time issue, the confidence ratings were recorded per answer (stage IIa) and reported insight (in stage IIb). As the research questions and hypotheses are formulated in comparison between 2D and 3D within stage and integration of findings across stages is additional for these aspects they do not influence the results directly.

Splitting up longer answers to single insights is closely related to the issue of rating the insights for complexity and plausibility. Splitting answers up in smaller insights leaves them potentially less complex and more plausible while longer insights are potentially more complex and less plausible. While insight ratings within researchers and between researchers are quite constant (section 2.6.2.2) there are some differences between researchers about the interpretation of what constitutes an insight. There is no easy way to solve this problem. North (2006) recommends using domain experts for the evaluation and coding of the insights. In this study it was not feasible to use domain experts as only half of the data sets of stages IIa and IIb were real. The other half was spatially transformed to have a second study setting (section 2.5.2.3). While this approach yielded valuable results in regard of detecting influences of data sets and settings (section 4.1.5) it removed the real relationship between data and landscape which a domain expert would be able to interpret more valuably than other coders.

Insight-based visualisation evaluation is a relatively new method introduce by Saraiya et al. (2005). Other researchers have discussed this approach (North 2006) or taken it up for their visualisation evaluations (Rester et al. 2007). This research has shown that it is a valuable approach even though some difficulties in application (e.g. exact definition of an insight) exist. The method will certainly gain from further usage and the combination of experiences to refine it. In addition, it may be worth considering the integration of approaches from related disciplines, for example, the measures of interestingness which are used in data mining (Geng & Hamilton 2006).

Implementation issues:
In stage IIa participation in the experiment (including instructions and training settings) was completely in a distance online mode. This has the advantage that a larger group of informed users could participate (section 2.5.2.4). Participation in all other stages took place in a face-to-face set-

ting where participants worked on a computer looking at the visualisations and completing digital questionnaires. Comparing the comments from all stages (section 3.6) shows that participants may have felt left alone with their problems and questions in stage IIa even though they could send an email if they encountered problems. In the face-to-face settings of stage I and IIb not many questions needed answering but being able to potentially ask questions may have given the participants a better feeling. In another context, Ocker & Fjermestad (1998) report that participants performed best with a mix of asynchronous and face-to-face work. The positive aspect of this problem is that participants in the online stage IIa settings commented in more detail about their experience with the visualisations than did stage I and stage IIb participants. Thus, the evaluation of comments may have gained even though some participants repeated comments. Comparing the number of comments within the task answers in stage IIa about not understanding aspects of the task (e.g. "what do the bars exactly mean", section 3.6.2) to the number of questions which needed answering during stage IIb it seems that participants listen more carefully to oral introductions compared to reading written instructions or they may have skipped instructions completely (Andrienko, Andrienko, Voss, Bernardo, Hipolito & Kretchmer 2002).

The 2D and 3D visualisations tested in this research are computer based. Thus, problems with hardware, software and operating systems may occur. While in real world settings such problems need to be solved, in research settings it often hinders participation (e.g. 14 potential participants in stage IIa, section 2.5.2.4, or three potential participants in stage IIb, section 2.5.3.4, could not participate) and potentially introduces a slight bias as only participants not encountering problems can take part.

As already discussed above, the participants of stage IIa spent more time than anticipated with the questionnaire (which included some training with the navigation in the 2D visualisation as well as in the 3D visualisation). While some participants did not participate because of time reasons, one participant only answered half of the questionnaire. Other participants may have started to hurry towards the end of the questionnaire as it took them more time than was indicated in the invitation to participate. There is a slight trend visible that confidence ratings are more often forgotten towards the end of the questionnaire indicating that participant indeed may have started to hurry or worked less concentratedly. However, employing balanced within-subject designs minimises the influence of this aspect on the results. The data is not corrected for participants who paused during answering questions. Some of the very long task completion times may be explained with participants taking a break. However, displaying the task completion times as boxplots of quartiles and median is robust against extreme values. In regard to time, the data experts informally reported that they thought their effort (4-5 hours) about right. None of the data experts would have liked to spend more time and they might have declined the invitation to participate if it had taken longer.

4.2.4. A mixed methods approach to bridge between *in vitro* and *in vivo*

The sections above, especially section 4.2.2, illustrate the many research aspects (findings and methodological design) that have benefited from either the sequential ordering of research stages and/or the possibility to evaluate and relate findings across the research stages. While the design and implementation of each research stage has strengths but also weaknesses (e.g. the experiment in stage I was very controlled but far from employing real world data set and tasks or the case studies of stage III include all the context information but many aspects cannot be controlled for and only few experts can participate) the re-testing of findings in further stages and/or relating findings across the stages has helped overcome or at least mitigate them. Thus, the mixed methods concept of ”complementary strengths and nonoverlapping weaknesses” (Johnson & Onwuegbuzie 2004, p. 18) proved to be helpful. The hypothesis that the main differences between ’in vitro’ and ’in vivo’ research are the differences in data and task complexity, number of relevant users and context (tacit knowledge of users and also the number of controlled for or allowed influencing factors) as illustrated in figure 2.1 is refined from the research results and experiences. The number of relevant users is certainly much larger in controlled (psychophysical) ’in vitro’ studies than in ’in vivo’ applied settings. However, the data and task complexity which was designed to increase from stage I to stage IIb was found to be potentially less complex in stage III. While data is often multivariate in real world settings this is not an imperative characteristic. Sometimes univariate data sets also have to be analysed. Additionally, in stage IIa it was found that, while still using univariate data sets, the specific characteristics of each data set influences the results. For task complexity, it was also found that depending on, for example, the data sets or the research aims the tasks can be of elementary and/or synoptic nature in real world applied settings. Instead of necessarily being more complex they may rather cover a wider range, clearly influenced by context. However, it was nevertheless valuable to design the experimental stages along increasing data and task complexity as it allowed a range of elementary and synoptic tasks to be tested and the detection of, for example, the difficulty of participants to relate data to location and altitude simultaneously, which is explicitly a synoptic task. The findings of stage III then challenge and thus also validate the findings from earlier stages.

The positions of the research stages along the bridge from ’in vivo’ to ’in vitro’ research methods were guided by increasing data and task complexity based on the hypothesis of the major changes between the two research continuum ends as discussed above. In reflection, it seems that stage I was neither really ’in vitro’ nor was stage III really ’in vivo’. An ’in vitro’ psychophysical controlled experiment in stage I would have required an even stricter control of influences and a much larger number of participants. The results from such as study could be evaluated for varying participants characteristics such as gender or spatial abilities. Using a group of informed participants, which are thought to have specific characteristics, offsets stage I slightly toward the applied end of the bridge. That makes sense in the overall research design according to the proposed research aims. However, it may have been valuable to be able to better control the influence of navigation. Knowing the kind of navigation and the viewpoint which finally led to the bar comparison judgement might tell us more about potential difficulties in judging bars in virtual environments even though the general result is that bar length judgements are possible in virtual environments. On the other hand, stage III was found to not really be ’in vivo’. The process as documented in the case study protocol with gathering data and context information before creating visualisations for discussion with the data experts may be closer to a user centred design and evaluation approach (Gabbard, Hix & Swan 1999, Robinson, Chen, Lengerich, Meyer & MacEachren 2005) than of testing and evaluating 3D visualisations ’in vivo’. The data experts suggested further improvements of the visualisation in regard to the current implementation and future uses of quantitative data displays within virtual environments and did not really work with the visualisation and their data sets. So, while the bridge yielded valuable results it could go on further to the left (’in vitro’) as well as further to the right (’in vivo’). Shneiderman & Plaisant (2006) suggest using Multi-dimensional In-depth Long-term Case Studies (MILCs) to evaluate information visualisations with domain experts in real world settings. A similar approach may be useful for the further evaluation of the appropriateness of quantitative data displays in virtual environments for the use in real world settings. A close long-term relationship between visualisation researcher and data experts may be needed to allow for such a study. It would need to take place when the potential is there, for example, the data expert has a new data set, research questions for the exploration of the data set, and when time and motivation to conduct the study is available on both sides. The data expert may also have to agree to first evaluate the new data sets with the new visualisation technique before evaluating with traditional methods to eliminate the confounding factor of familiarity with the data set. Back to this research, in the design phase the subdivision of stage II in stages IIa and IIb was decided to narrow the gap between the research stages. In reflection, this decision was right. The results from stage IIa and IIb and also from the comparison of these two sub-stages yielded detailed findings in regard to influences such as data sets, settings, task definition and complexity which would have been more difficult if not impossible to elicit from a single stage II.

In this research the results of the quantitative and qualitative analysis are reported separately (chapter 3) and later integrated in discussion (this chapter) as this is a straightforward way to improve the readability of the report and also facilitate references to sections with additional details. Sometimes (e.g. Bazeley 2004) the coding of qualitative text and subsequent statistical analysis of the number of coded text fragments, as done in this research, is referred to as integrative quantitative and qualitative analysis. However, it is rather a transformation of qualitative data into quantitative data (quantify qualitative findings, Tashakkori & Teddlie 1998) which may facilitate easier comparison and integration of the results (Creswell, Plano Clark & Garrett 2008). Using 'thick matrix displays' for the analysis of the stage III data sets was valuable. It allowed details of each case to be seen at the same time as comparing issues across cases was possible. The decision whether to report 'thick matrix displays' of case data as tables in this report (section 3.5.2) is useful is left to the reader. It is supposed to be valuable as it makes the conclusions more transparent and may also help readers to experience the detail of the collected data and to compare the case characteristics and information to their own applications.

The main purposes for a mixed methods approach were complementarity (to elaborate and enrich findings from one method with results from another), expansion (aiming for breadth) and triangulation (seeking convergence of findings) (Greene et al. 1989). As illustrated in this discussion chapter, in the first part in regard to contents (section 4.1) and in this second part in regard to methodological aspects, complementarity, expansion (especially through stage III) and triangulation is achieved. Thus, mixing quantitative and qualitative methods is not only employed to explain numbers with text and add statistical significance to text with numbers but to gain a more holistic view of the use of quantitative data displays in virtual environments in a range of settings. Employing a series of research stages is probably less profound than in-depth studies of single aspects but more holistic and may yield applicable results faster. The lesser depth in one stage can be mitigated by re-evaluating the results in further stages even though the focus of these stages may be slightly different. Meng (2003, p.1) stated some years ago that "the development of cartographic theories and methods lags far behind the technical evolutions." Thus, holistic approaches may be needed nowadays and in the future to be able for theory to keep up with the technological developments or at least to provide some guidelines on how the fast evolving technologies might be appropriately utilised for data analysis. The amount of data to be analysed and thus the number of developed or adapted visualisation techniques might rather increase than decrease in the future.

5. Conclusions and outlook

< Chapter 4. Discussion

Overall the aims and research objectives of this research are achieved (5.1). This conclusion is based on the discussion of results within and across the research stages. Understanding of the appropriateness of quantitative data displays within virtual environments is increased by gaining a holistic view of the issue in different experimental and applied settings with varyingly complex data and tasks. Nevertheless, the details, such as data graphic complexity or the influence of data sets and settings, are taken into account.
The conclusions on the visualisations are (5.2.1):

- 2D bars or bar charts with reference frames on billboards are appropriate for the display of quantitative data in desktop-based 3D virtual environments for elementary and synoptic tasks in a range of settings and applications.
- From comparing the 2D and 3D visualisations it is concluded that the 3D display of the landscape and setting does neither help nor hinder the interpretation of altitude differences and landform. Similarly, the 2D display of the landscape and setting does neither help nor hinder the interpretation of positional aspects such as distribution.
- Different data sets, and the settings they belong to, strongly influence the results of the appropriateness measures in the 2D and the 3D visualisation. Some data sets are by trend better analysed within 3D visualisations, others in 2D visualisations, and a third group of data sets gains from analysis in 2D and 3D visualisations in regard to specific appropriateness measures.
- Visual analysis of data sets, especially in relation to altitude and landform, is not common but people can do it if a task requires it.
- Displaying all data which is relevant for the analysis concurrently is imperative in applied settings.
- Navigation (viewpoint change) is a cost but imperative for the analysis of quantitative data within virtual environments.

The conclusions on the methods are (5.2.2):

- It is appropriate to use a sequential mixed methods research approach with research stages guided by increasingly complex data sets and tasks to gain a holistic understanding of a visualisation technique.
- Task definitions based on the functional view of data and tasks (Andrienko & Andrienko 2006) are appropriate. The functional view facilitates in the first place, and supports, a precise definition of tasks in relation to the interesting dimensions (e.g. altitude and location) of the visualisation evaluation (including the dissociation of data and navigation/interaction tasks).
- Insight evaluation is a valuable technique for analysing the outcomes of explorative visual data analysis.
- It is appropriate to use a group of informed participants to evaluate a new visualisation technique experimentally with limited influences of context information as done in stages I, IIa and IIb.

Recommendations on the use of quantitative data displays in virtual environments are made (5.3.1). Issues such as data display type, reference frames, data display preparation and characteristics of data sets which may benefit from visual analysis in virtual environments, are covered. Additionally, recommendations on the use of evaluation methods are made (5.3.2). Issues such as mixing methods, task definition, answering tasks in comparison to reporting insights, providing training and implementation considerations, are covered.
The results of this research may influence future evaluations of visualisation techniques (5.4). Experimental results, as from stages I, IIa and IIb, may serve as input for future visualisation design and evaluation processes. Additionally, based on the results it seems important to assure participants about their visual analysis capabilities. The process of identifying specific data set characteristics that aid the decision in what type of visualisation the data are best analysed, is started with the results of this research.
Before giving an outlook on potential future uses and applications of quantitative data displays in virtual environments (5.6), four issues which need further research are identified (5.5): labelling in virtual environments, interaction with the display and the data, combinations of 2D and 3D displays, and the use of different data graphics.

Table of contents 2

5.1. Revisiting research aims and objectives

Based on the data analysis of each stage (chapter 3) and the discussion of the results within and between research stages (chapter 4) it is concluded that overall the aims and research objectives (table 5.1) are achieved. Only the main hypothesis of stage III could not be evaluated on the basis of the collected data as detailed in sections 3.5.3 and 4.2.1.

aims	✔ To increase understanding of the use of desktop-based 3D virtual environments with a focus on the graphical representation of quantitative data through abstract symbols or graphics. ✔ To relate perceptual/cognitive studies and case studies in specific application areas by evaluating the use of desktop-based 3D virtual environments for analysing quantitative data in a range of experimental and applied settings.		
research objectives	✔ Are methods of quantitative data representation derived from literature about 2D representation methods appropriate in desktop-based virtual environments in an experimental context?	✔ Are these methods appropriate for more complex visualisation tasks involving landform?	✔ Are these methods appropriate in applied scenarios with real tasks and data experts for the visualisation of quantitative data in relation to landform?
	stage I	**stage IIa & IIb**	**stage III**
main hypotheses	✔ Typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative values in virtual environments when using appropriate representation methods.	✔ Typical users are able to relate multivariate data representations displayed in virtual environments. ✔ Typical users are able to relate appropriate representations of quantitative data to landform.	✳ Data and task experts gain insights into their data sets when appropriately displayed in virtual environments depicting the landscape to which the data relates.

Table 5.1.: Overview of explored and tested aims, research objectives and main hypotheses (✔ = accepted/yes, ✘ = rejected/no, ✳ = not answered)

While table 5.1 shows the overall view of evaluated research aims, objectives and hypotheses which are mostly accepted or positively answered, a number of underlying factors could be detected and evaluated during the research process. Generally, the understanding of using virtual environments for the visual analysis of quantitative data has been largely increased as the evaluation of different data sets, settings, tasks and contextual influences in the various stages allowed getting a holistic view of the issue. But the design nevertheless also permitted attending to specific aspects and details of representing quantitative data in virtual environments. The following sections explain the conclusions from this research (including scope and limitation), give recommendations on the use of the visualisations and the methodological framework and offer starting points for further research that builds upon these findings and recommendations.

5.2. Conclusions

5.2.1. Concluding on the visualisations

5.2.1.1. Data graphics

Conclusion: 2D bars or bar charts with reference frames on billboards are appropriate for the display of quantitative data in desktop-based 3D virtual environments for elementary and synoptic tasks in a range of settings and applications.

This conclusion is based on the design of this type of data graphic from knowledge of 2D cartography and perception research (section 2.3.1) and the results of testing this type of data graphic in all four research stages with data and tasks of varying complexity and with informed users (section 2.4) and data experts. Typical users are able to separate depth cues and distortions introduced by perspective viewing from absolute value changes in the representations of quantitative data in virtual environments when represented as 2D bars or bar charts on billboards.

Scope and limitation:
Data graphic (sections 4.1.1 and 4.1.2): This evaluation focusses on quantitative data displays as 2D bars on billboards as derived from literature. In another study (Bleisch accepted) this type of symbology is compared to 2D circles and 3D bars based on performance with elementary tasks and proved to be the most efficient and effective symbol type of the tested ones. However, no other types of data displays were tested with more complex data and tasks in this research. From an aesthetic point of view, it may be argued that 2D bars on billboards do not integrate too well in the otherwise three-dimensional landscape. Nevertheless, data values mapped onto bar length on billboards can be compared and analysed in virtual environments. Estimates of bar length comparisons (percentage of the smaller compared to the taller bar) in virtual environments can be compared to the true comparison values. The mean of these differences between estimated and exact bar length comparison values is 1.64. Thus, the smaller bar is slightly overestimated in comparison to the taller bar. Participants remarked that, generally, small bars are more difficult to compare than tall bars. Careful consideration of existing 2D cartographic knowledge together with knowledge about the human perception of 3D displays is valuable for the design of data displays in virtual environments.

Reference frames (section 4.1.1): Reference frames are helpful for the comparison of bar displays in virtual environments. Exact estimation of bar lengths is only possible with reference frames.

Additionally, they help grouping the data values belonging together and allow localisation of the area a data graphic belongs to. However, as may be expected, when data sets are more dense and/or data visualisations are looked at from a zoomed out view the reference frames crowd the display.

Data graphic complexity (section 4.1.1): Displaying multivariate data sets by showing several variables on one data graphic as combination of bars which are either stacked and/or positioned beside each other is possible. The participants identified a complexity threshold which they do not like to see crossed. A maximum of 3-4 variables should be shown on the same data graphic. Stacked bars are preferred over bars beside each other as they crowd the display less. However, they may not be aware that inter value comparison is more difficult with stacked bars.

5.2.1.2. Relation of data and landscape

> Conclusion: From comparing the 2D and 3D visualisations it is concluded that the 3D display of the landscape and setting does neither help nor hinder the interpretation of altitude differences and landform. Similarly, the 2D display of the landscape and setting does neither help nor hinder the interpretation of positional aspects such as distribution.

This conclusion is based on the comparison of appropriateness measures between 2D and 3D visualisations in the stages I, IIa and IIb (section 4.1.3.2). No significant or relevant differences between the 2D and 3D visualisations as tested in this study are found. Thus, the proposed advantage of 3D visualisations ”providing a familiar view of the world” (Shepherd 2008, p. 202) and thus helping the interpretation of altitude and landform is not reflected in the appropriateness measures used in this research (cf. section 5.2.1.4, visualisation preference).

Scope and limitation:
Comparison of 2D and 3D visualisations (section 4.1.3): The analysis of data values in relation to location and altitude is compared between 2D and 3D visualisations in stages I, IIa and IIb of this research. The 2D visualisation serves as a benchmark as such types of displays based on maps are traditionally used to analyse data sets in relation to the landscape. However, these 2D and 3D displays are not compared against other visualisations of the same data sets. Additionally, while comparing new methods to established ones is helpful, as measures can be baselined, it may not do justice the potential of the new visualisation technique (Greenberg & Buxton 2008). The high popularity of virtual globes and 3D visualisations and the strong feelings of participants for or against it may indicate that there is more to this visualisation type than can be evaluated in (experimental) comparisons.

5.2.1.3. Data sets and settings

> Conclusion: Different data sets, and the settings they belong to, strongly influence the results of the appropriateness measures in the 2D and the 3D visualisation. Some data sets are by trend better analysed within 3D visualisations, others in 2D visualisations, and a third group of data sets gains from analysis in 2D and 3D visualisations in regard to specific appropriateness measures.

This conclusion is based on the evaluation of eight different data sets and two settings in stage IIa, two different data sets in the same two settings in stage IIb and observations and statements from the case studies of stage III.

Scope and limitation:
Data sets, and the settings they belong to, influence the outcomes but it is difficult to say how exactly. The recommendations made (sections 4.1.5 and below 5.3.1, characteristics of data sets) are based on the evidence collected from the use of different data sets in this research. Generally, a less complex setting in terms of variation of topography, land-cover and data set tends to yield less complex insights of higher plausibility and with higher confidence ratings. Settings with a readily visible combination of upper/open vs. lower/wooded areas leads to a more frequent use of combined references in insights. Most data sets employed in this research (mostly point data sets) were spatially and/or thematically aggregated before displaying them as bars or bar charts in 2D and 3D. Aggregation results in loss of detail while aiming for improved clarity of the display. However, different aggregations (e.g. in differently sized grids in stage III, case 'Deer SNP') may yield different displays. Thus, aggregation potentially influences the results but this issue is not researched here as the comparison between 2D and 3D is not affected.

5.2.1.4. Ancillary conclusions

Visual analysis of data

> Conclusion: Visual analysis of data sets, especially in relation to altitude and landform, is not common but people can do it if a task requires it.

This conclusion is based on the evaluation of participant comments in all stages (section 4.1.7). Additionally, participants in stage IIb not guided by tasks (in comparison to stage IIa) resorted to evaluating and reporting insights with reference to location much more often than reporting insights with references to altitude or a combination of location and altitude (section 4.1.6). However, participants can do analysis of the relationship between data and landform as shown in stage IIa. There, guided by the tasks, participants gave answers relating data and landform but report that the tasks are difficult to complete.

Visualisation completeness

> Conclusion: Displaying all data which is relevant for the analysis concurrently is imperative in applied settings.

This conclusion is based on the reported and observed visualisation usefulness in stage III (section 4.1.7). Useful visualisations in applied settings are able to display all relevant data concurrently.

Visualisation preference

> Conclusion: Participant's preference for the 3D visualisation has a quantitative effect by resulting in more reported insights (in 2D and 3D) rather than a qualitative effect in comparison to participants stating preference for the 2D visualisation.

This conclusion is based on weak evidence from the evaluation of appropriateness measures in stage IIa for three participants preferring the 2D visualisation and seven participants preferring the 3D visualisation (section 4.1.3.2). As the differences between the 2D and 3D displays as evaluated in this research are minimal, participants could use either of the two visualisation types. However, appropriateness of the display may vary depending on the data set to be analysed (section 5.2.1.3).

Navigation

Conclusion: Navigation (viewpoint change) is a cost but imperative for the analysis of quantitative data within virtual environments.

This conclusion is based on the evaluation of participant comments in all research stages. Navigation is an influencing factor which is not tested for but which is kept constant by using standard interaction and navigation functionality in the 2D SVG displays and the 3D visualisations based on Google Earth (sections 2.2 and 2.3.2). The navigation in both the 2D and the 3D visualisations takes time and is remarked upon as being unfamiliar (section 4.1.3). Nevertheless, the navigation in the 3D visualisation is commented upon as being essential to overcome data graphic overlaps and/or landscape occlusion, relate the data to landform (including localisation of the data graphics) or to get a different view of the data and/or topographic features. Thus, navigation is a cost but one users can benefit from.

5.2.2. Concluding on the methods

Mixed methods approach

Conclusion: It is appropriate to use a sequential mixed methods research approach with research stages guided by increasingly complex data sets and tasks to gain a holistic understanding of a visualisation technique.

This conclusion is based on the analysis of each research stage separately and the combination of the findings (chapters 3 and 4). The research aims and objectives are achieved. The results give a holistic but also detailed view of the appropriateness of displays of quantitative data within virtual environments in a range of settings including varyingly complex data sets and tasks. None of the research stages alone could give the answers looked for. The results of the various stages strengthen but at the same time question the findings. For example, the reference frames which are helpful for simple tasks in stage I crowd the display in more dense setting, defining specific tasks in stage IIa guide the users to relate data to landform but restrict their freedom of exploration or data analysis of a specific data set, which is done by the informed participants without question in experimental settings, is more difficult in stage III where data experts are (over-) familiar with the data sets and emphasise the importance of displaying all relevant information of a data set.

It is argued that the approach of using different research methodologies along a continuum of increasing data and task complexity to evaluate a visualisation technique 'in vitro' and 'in vivo' benefits from each research stage to build an appropriate and valuable 'bridge' of knowledge. The knowledge of this research design should be applicable to other evaluations of research techniques where researching a single stage, be it 'in vitro' or rather 'in vivo' does not allow a holistic view to be gained. Researching display techniques experimentally requires research in more applied settings to validate the results while researching visualisation techniques in applied settings are much less generalisable without the possibility of combining the findings with data and results of other stages with more participants as done in this research. However, implementing and evaluating four different research stages is rather costly. Section 5.3.2 below recommends a pragmatic approach which may allow the most important characteristics of a new visualisation technique to be gathered through two research stages only.

Scope and limitation: This research was neither designed as a usability study (mainly concerned with interaction, navigation and the interface, e.g. Stanney, Mollaghasemi, Reeves, Breaux & Graeber 2003) nor as a user-centered design and evaluation approach (Gabbard et al. 1999). Fuhrmann, Ahonen-Rainio, Edsall, Fabrikant, Koua, Tobon, Ware & Wilson (2005, p. 555) state that the "often system-focused HCI methods might not distinguish between useful and usable". This research is concerned with 'useful' or 'appropriate' (section 2.1). The focus lies on the evaluation of the visual analysis of quantitative data in relation to the landscape it belongs to by employing a visualisation technique (quantitative data graphics within desktop-based 3D virtual environments) which is hypothesised to support this type of analysis. The aspects of varyingly complex data sets (uni- and multivariate) and varyingly complex tasks (including varying amounts of context information) are used to be able to evaluate appropriateness measures in a range of settings and applications. The aspect 'useable' (e.g. navigation) is tried to be kept standard and/or constant and is not in the focus of the research. However, it is acknowledged that navigation and interaction are important factors which require consideration and research.

The experiences and findings of stage III have shown that the design did not really allow the continuation of the methodological bridge 'in vivo'. Several factors such as the data experts being very familiar with the data sets, having no immediate interest in the (re-)analysis of their data sets or limited time, influenced this. In some parts, stage III rather resembled the first steps of a visualisation design process (Gabbard et al. 1999) by gathering data and typical tasks from the application area, creating visualisations and discussing the (prototype) visualisations with the data experts (cf. section 5.4).

Tasks

> Conclusion: Task definitions based on the functional view of data and tasks (Andrienko & Andrienko 2006) are appropriate. The functional view facilitates in the first place, and supports, a precise definition of tasks in relation to the dimensions of interest (e.g. altitude and location) of the visualisation evaluation (including the dissociation of data and navigation/interaction tasks).

This conclusion is based on the evaluation of the results of the research stages I, and especially IIa and IIb where the functional view of data and tasks allowed tasks and insights to be separated according to whether they referred to location, altitude or a combination of location and altitude. This separation allowed the hypotheses to be answered. The reported typical research questions of stage III could be analysed based on the functional task view and compared to the tasks of earlier research stages. Additionally, the functional task view allowed the dissociation of data and navigation/interaction tasks what is often difficult when using other task taxonomies (section 4.1.4).

Scope and limitation: Reports about difficult to answer tasks are related to the fact that some tasks did not have a clear answer in some data sets (section 4.1.4) and, additionally, participants are not used to data analysis in regard to altitude and landform (section 5.2.1.4). However, the wording of tasks does influence at least the wording of the answers and presumably also the content (section 4.1.4). While this is not generally bad (answers answering the task are normally looked for) it may nevertheless be sensible to consider this influence when setting up tasks (cf. section 5.3.2, answering tasks vs. reporting insights). Comparing the tasks derived from the typical research questions reported in stage III shows that the tasks employed in earlier stages are representative of real world tasks even though they may not cover the whole breadth of possible real world tasks which are dependent on the context, data sets and application areas.

Insight evaluation

Conclusion: Insight evaluation is a valuable technique for analysing the outcomes of explorative visual data analysis.

This conclusion is based on the results of evaluating insights in stages IIa and IIb which allowed the research questions to be answered (sections 3.2 and 3.3). North (2006, p. 6) proposes that "the purpose of visualisation is insight." In this research, insight is collected in two different ways. In stage IIa insights are evaluated within answers to given tasks. In stage IIb insights within insight reports are evaluated. In both situations, the method allowed the research questions of each stage to be answered but, additionally, it facilitated the integration of results between the two research stages.

Scope and limitation: It is acknowledged that insight evaluation is a somewhat subjective method. However, with more complex data sets and tasks, as for example in explorative data analysis, there are no clear cut wrong or right answers that can be evaluated (in difference to the tasks in stage I where correct answers are available). The objectivity of the methods is maintained through strict coding practices (section 2.6.2) and stability and reproducibility is checked through evaluating intra-coder and inter-coder reliability.

Informed participants

Conclusion: It is appropriate to use a group of informed participants to evaluate a new visualisation technique experimentally with limited influences of context information as done in stages I, IIa and IIb.

This conclusion is based on the triangulation of findings of the first three stages with data experts statements from stage III. Many aspects in regard to data graphics, task complexity and the 3D virtual environment in general could be detected in the earlier stages (section 4.2.2). However, some aspects of using 3D visualisations in applied settings were only obtained through the case studies with data experts (e.g. importance of visualisation completeness or detailed descriptions of potential applications, section 3.5.2).

Scope and limitation: The informed users who participated in this study mostly showed similar characteristics in education and experience in visual analysis of geodata as is expected from the data experts in the selected case studies. These characteristics need some consideration even though they cannot be controlled for in every detail. Baselining the data experts with tasks the other participants do (stage I elementary tasks) is valuable as it helps the comparison between informed participants and data experts (section 4.2.2).

Motivation may play a significant role when participants take part in a study. As shown above (section 5.2.1.4), visualisation preference does have some influence on the results. Purpose in task and goals is important for the participation and thus the obtained results (Marsh 2007). While informed participants without a genuine interest in exploring the data may get bored and/or report trivial findings (e.g. of low complexity) data experts, on the other hand, may be (over-)familiar with their data sets and context and thus be hindered in data exploration (cf. section 5.3.2, answering tasks vs. reporting insights).

5.3. Recommendations

5.3.1. Data displays in virtual environments

The following recommendations for quantitative data displays within desktop-based 3D virtual environments are made based on the findings and experiences of this research. While some of them are based on the research results others come from the experience in preparing the data visualisations for the various stages of this research. Further, some proposals for consideration are included which may need further research but seem plausible based on the experiences of this research.

From the results it is concluded that the appropriateness of quantitative data displays in 3D virtual environments is dependent on data set and setting characteristics. This should be considered when applying the following design recommendations.

Data display: Use the length of 2D bars on billboards to represent varying quantitative data values in virtual environments such as Google Earth which are viewed on a desktop screen. When displaying multivariate data sets try to avoid displaying more than 3-4 variables at once and use stacked bars rather than arranging several bars beside each other in bar chart displays. These arrangements shall ensure that data graphic and/or landscape overlap and occlusion is as minimal as possible.

Using stacked bars, however, biases bar comparison. The bottom part of a stacked bar is most easily compared. Interactive displays could accommodate this issue by allowing the order of the stacked bars to be changed. Thus, the variable which is to be compared is displayed at the bottom of the stacked bar. But this was neither implemented nor tested in the course of this research.

Based on first experiments with elementary tasks, 3D bars with reference frames are similarly effective as 2D bars on billboards. However, they have not been tested with more complex tasks or multivariate data sets. The issue of crowding references frames in more dense displays may become even worse with 3D bars as they consist of many more reference frame lines.

Reference frames: Use reference frames for your bars or bar charts in virtual environments showing either the maximum data value displayed or potentially a reference value (e.g. 100%). Reference frames allow more accurate bar length judgement and thus also bar length comparison. In addition, they are helpful for localising the data values in the landscape. In this research, black reference frames showing the maximum data value were used. One data expert tentatively suggested to use white reference frames where the background is dark.

Reference frames crowd the display when data density is high and/or the data is viewed from a zoomed out view (overview) making them less useful. You may want to consider adding interaction to your data display allowing the reference frames to be switched on/off as asked for by some participants. However, the influences of such interaction on the appropriateness of the visualisation is not tested in this research (section 5.5).

Data display preparation: Beside this research a process for the preparation of data displays for virtual environments was developed and implemented (section 2.3.2). The process was thoroughly tested by creating all data displays used in this research and it proved very helpful. The following aspects need consideration beside automating the data display preparation process.

- Use spatial and thematic aggregation carefully and consider the implications on the data display. From other contexts (e.g. classification for thematic mapping, Slocum et al. 2005) it is known that different data aggregations create different visualisations and thus different impressions of the data. Andrienko & Andrienko (2006) provide guidance on data aggregation.
- Scale your data displays appropriately. This is generally dependent on the size of the area a data set belongs to (make the data displays visible while at the same time avoiding overlap). Additionally, the ratio of the smallest to the largest bars should be optimised as small bars are less visible and more difficult to compare (e.g. by considering a different aggregation or by displaying sub-data sets).
- Provide a correct legend explaining all variables displayed (not yet natively supported by the data graphic preparation process referred to above).

Characteristics of data sets which appear to benefit from visual analysis in 3D visualisations: The research findings suggest that 3D visualisations are more appropriate for the analysis of more complex data sets, as characterised by being denser, having less clustered data points but larger, visually prominent data values covering a potentially larger area with varied topography and landcover (cf. section 4.1.5 for the complete discussion and examples). However, further research may want to evaluate the results if 3D visualisations contain larger, visually prominent values which are not important for task answers but may nevertheless lead the participants to answer the tasks as they seem to do in this research. In regard to specific appropriateness measures, some data sets gain from an analysis in the 2D and the 3D visualisation (section 5.5).

5.3.2. Evaluation methods

The following recommendations on the use of evaluation methods are made based on the findings and experiences of this research. While some of them are based on the research results others come from the experience in preparing and analysing the various stages of this research. Further, some proposals for consideration are included which may need further research but seem plausible based on the experiences of this research.

Mixing methods and bridging research stages: Use different methods if they are justified based on the research aims and objectives. As exemplified in this research, mixing methods sequentially has benefits as, for example, being able to adjust methods slightly from one stage to another based on the earlier results (e.g. seven tasks in stage IIa vs. one task in stage IIb). However, at the same time the changes in methods may not completely be separable from changes in dimension of interest in the evaluation. The complementary methods and also the data analysis methods (e.g. replication strategy) should mitigate or overcome such problems. Importantly, using different research stages allowed the findings to be compared forward and backward along the stages to build a valuable bridge of knowledge. This is not only dependent on using various methods but also on the identification of suitable stages in regard to the hypothesised differences between the stages and the defined research aims and objectives. The criteria used to identify the different research stages, increasingly complex data and tasks in this research, need consideration and justification.

Consider that using various methods needs theoretical and practical knowledge for the design and application of each method. Familiarisation with a new research and/or analysis method can take up a considerable amount of time.

In section 5.2.2 it is concluded that it is appropriate to use a sequential mixed methods approach with different research stages to gain a holistic understanding of a visualisation technique. However, implementing and evaluating four different research stages is rather costly in terms of time, methodological knowledge and number of participants. From the experience and results of this research it is suggested that two research stages may suffice to gain an understanding of a visualisation technique. A first stage would be similar to stage I of this research. A visualisation method is derived and to some degree justified from existing knowledge. This method is then tested with informed users in a controlled experiment. The experiment could include a few more than two values to compare and potentially also another simple task (cf. section 5.4 on validating participants abilities). A second stage would then test with more complex (multivariate) data sets and tasks (a combination of stages IIa and IIb of this research). As tasks influence the performance and answers (cf. paragraph on answering tasks vs. reporting insights below) a between-subject design in regard to different tasks and a within-subject design in regard to data sets (use more than one) and maybe also visualisation types may be employed. The results of these two stages may give a good understanding of the visualisation method and could valuable feed into a visualisation design process and/or later evaluation 'in vivo'. To close the gaps between visualisation researchers and application experts (van Wijk 2006), design processes and/or further evaluation of experimental results 'in vivo' is imperative. However, this may not necessarily need to take place in the form of stage III of this research (cf. section 5.4).

Task de nition: Employ Andrienko & Andrienko's (2006) functional data and task view which allows precise tasks to be defined which vary in the dimension of interest (e.g. location and/or altitude as reference set) for the evaluation. Other task taxonomies often combine data and interaction tasks and do not allow exact categorisation of tasks (section 2.2). But this is required to be able to separate between the evaluation of either 'useful' or 'useable'. The definition of the task influences the answers which can be obtained (sections 4.1.3.2 and 4.1.4). Comprehensible task and instruction wording is supported by applying the TAP paradigm (topic, applicability and perspective, Foddy 1994). Not only task definition and wording, as found in this research, but also instructions can influence experimental results (reported for instructions on distance-similarity relationship judgements by Fabrikant & Montello 2008).

Answering tasks vs. reporting insights: This is not a recommendation but rather an offer of alternative possibilities with distinct advantages and disadvantages. Based on the experiences and results of stages IIa (collecting answers to seven tasks) and stage IIb (reporting insights based on one explorative task) the following characteristics of each method are found. Participants guided by seven tasks they had to answer per data set tried to answer whenever possible even though some tasks did not have an explicit answer in some data sets (section 4.1.6). They related the data to landscape (altitude and/or location) as asked for in the tasks but reported that they find some tasks difficult to answer. Participants reporting insights based on a single explorative task were more free in their exploration which is visible in the additional time they take to report an insight (1.7 times the time in 3D, but this is also dependent on the more complex data sets which have to be evaluated) and the higher confidence they have in their findings. However, they may only report the findings they are confident in compared to the other participants trying to complete all of the tasks even if they find it difficult. Participants reporting insights show clearly that they are not used to analysing the relation of data to landform by significantly more often reporting location based relationships between data and landscape. Additionally, some participants seem to get bored with the exploration of the visualisation based on a single tasks and did less comprehensive analysis of the data. This results in fewer or more trivial insights being reported towards the end of a exploration phase.

So while the tasks provided guidance, which may be important especially for less experienced users, they also influence the results as (only) answers are provided. Insights into the data set which are not the answer to a task are left out. Participants reporting insights work closer to real applied settings by exploring the data sets based on an explorative (research) task. They have more freedom and are more confident in the findings they report but may get lost in exploration or get bored. To gain the best of both sides, a combination may be worth considering. For example, let the participants explore freely but occasionally (or when they stop recording insights regularly) provide them either with further questions and/or with input about the type of insight looked for (e.g. relation of data to landscape) to stop them of getting bored and providing them with some guidance. Evaluations with data experts are less at risk of this problem as data experts (often) have an intrinsic interest in the data exploration and thus may not get bored. But on the other hand they may already be familiar with the data set and thus be less motivated to (re-)analyse it.

For the evaluation of task accuracy it is imperative to define exact tasks with previously known correct answers that the given answers can be compared to (e.g. bar comparison in stage I, section 4.1.2). For the analysis of answers to explorative tasks insight evaluation is an appropriate method (section 5.2.2, Insight evaluation). However, strict coding practices have to be implemented and the evaluations need to be checked for reliability and validity.

Provision of training/guidance: Based on the conclusion that informed participants and also data experts in this research were not used to the (visual) analysis of data in relation to altitude and landform it is imperative to provide some training and/or guidance. As remarked above (paragraph Answering tasks vs. reporting insights), participants do analyse the relation between data and landscape when they are guided by specific tasks they need to answer. However, letting them explore more freely they rather report insights in relation to location. Thus, you may either want to train the participants with test settings and examples of the kind of information being sought before letting them explore or guide them with specific questions taking into account that only the questions will be answered and no further insights looked for. With a large enough number of participants both methods could be implemented concurrently using a between-subject design and be evaluated in comparison. The provision of training may also be helpful especially for participants not used to visual analysis of data sets, for example, because they use visualisations for communication purposes only (section 4.1.7).

Implementation: If possible and not against the research aims and objectives it should be aimed to experiment with participants in a face-to-face setting while, nevertheless, employing digital data collection means (e.g. a digital questionnaire to avoid potential digitalisation errors). In this research, participants tended to listen to/follow instructions more carefully in the face-to-face settings than in the distance online setting even though few questions needed to be answered (section 4.2.3). Face-to-face participants wrote fewer comments but seemed more at ease with the experiment and the tasks.

5.4. Implications

The results of this research show that generally displays of quantitative data as 2D bars on billboards in virtual environments perform similarly well as similar 2D map based representations. However, the results may vary for different application areas, data sets, tasks, users etc. The following sections give an overview and suggest how these varying influences may possibly be accounted for in practice.

The findings of this research, especially from stages I, IIa and IIb or results from more pragmatic two-stage visualisation evaluations (as suggested in section 5.3.2), could feed into a visualisation design process. In an adapted version of stage III the findings from the earlier stages could be validated and complemented. In this research, the aim of evaluating the 3D visualisations 'in vivo' was not completely achieved (section 4.2.4). Stage III resembled to some degree the beginning of a visualisation design process and could potentially be adapted to serve this purpose. In this research, in stage III information about the context and the data set(s) were collected, a (prototype) 3D visualisation with the data sets designed and it was tested and commented upon by the data experts. The data experts were expected to work with the visualisation and analyse their own data sets. They did not do so as much as anticipated. They explained their data sets within the virtual environment and then explored the possibilities the visualisation has to offer for future data analyses and commented upon problems and potential improvements. Such information could be used to improve the data displays within virtual environments generally but also especially for the requirements of the specific application areas. Final evaluation of the visualisation technique truly 'in vivo' should take place when the design has sufficiently improved as suggested in design processes (e.g. Gabbard et al. 1999). Methods for such an evaluation could include long-term studies (e.g. MILC studies, Shneiderman & Plaisant 2006). Thus, results and conclusions from experimental methods based on increasingly complex data sets and tasks evaluated by informed users (similarly to stages I, IIa and IIb in this study) could valuably inform visualisation design processes. This especially for new types of visualisations techniques (e.g. different data graphic or symbol types) which have not been extensively evaluated and basic knowledge about performance with various tasks and data sets is missing so far. However, it needs to be considered that promising new ideas can be crushed if they have to compete too early against established techniques (Greenberg & Buxton 2008). Additionally, design processes could also take up the data experts' (section 3.5.4) and participants' (section 3.6.3) comments about navigation and interaction in the visualisations. The detailed data collected in stage III of this research, reported in 'thick matrix displays' (section 3.5.2), may also serve as input for ideation and visualisation prototype building in application areas not considered in this research.

Another influence on the results comes from the users and their background. The author of this study works partly in a technical engineering environment where visualisations are often regarded as a mean of illustration and communication only. 'Real' data analysis is done by calculations, statistics and numbers. Some of the study participants and data experts also come from this environment which shows up occasionally in the comments. Implementing extensive design processes for visualisations are rarely an option in such environments. For the promotion of visualisations for data analysis and exploration purposes people need to see that they can do it and that it works for their data sets and applications. Both aspects are important, informed participants took part in this study and especially data sets and settings are found to be confounding factors. The following two paragraphs try to sketch how users could be informed about their visual analysis capabilities and how it may be possible to account for various data sets and settings.

Users: Generally, the participants of this research are a group of informed users (section 2.4) valuably evaluating the visualisations in the stages I, IIa and IIb. They are not a homogeneous group but they have similar characteristics. Even though explorative geovisualization is rather for single data experts than for a mass of average users, the individual differences of the informed participants are not detailed in the evaluations (section 4.2.2, Participants). Individual differences are difficult to account for in experimental settings which require a certain number of participants to make inferences. However, for individuals, who would like to use the visualisation technique tested here, it may be sensible to reassure him or her that they belong (or do not belong) to the group of people who can interpret data values in virtual landscapes even though such interpretation may not always feel easy (e.g. the relation of data and landscape which participants are not used to do).

Informing users about their capabilities in visually analysing quantitative data within virtual environments could take the form of a quick experiment. An application could provide simple data sets and elementary tasks which need to be answered, similar to stage I of this research. It may be sensible to complement the two elementary tasks of stage I ('lookup' and 'comparison') with the third elementary task 'relation seeking' (Andrienko & Andrienko 2006). However, a 'relation seeking' task would require a data set showing a few more than two single data values. For validation, a user would start the application, work through the tasks and submit the answers. The system would in return provide the user with the information about efficiency and effectiveness (task accuracy and task performance time). For the effect of baselining, the user should be enabled to compare his or her task performance to the average and/or range of other users' task performances. A much lower efficiency and/or effectiveness than average and/or a task performance outside the range of 50% of the other users' performances would indicate that the user's capabilities are potentially not sufficient for the analysis of quantitative data in virtual environments. Thus, the user may want to consider other means of analysing a data set. Informing users about their capabilities through efficiency and effectiveness measures in a quick experiment may assure them that they are able (or not able respectively) to do visual analysis in the tested display type and may make them more confident in their findings. Additionally, it may also help promoting visual data analysis as users get an indication about how efficient and effective it is, at least with simple tasks and data sets. But as the results of this research show, the simple data sets and tasks of stage I are an important basis of more complex data sets and tasks. The results of stage I about the use of 2D bar displays within virtual environments are not seriously questioned in later research stages. Quick user testing could also be implemented for different types of data displays (e.g. 3D bars). A user would thus be able to learn which display type works best for him or her and find out whether task performance is aligned with personal and/or aesthetic preferences.

Data sets: Validating if a visualisation technique is appropriate for a specific data set is more difficult. Based on the different data sets analysed in stages IIa, IIb and III a tentative proposition is made which data set characteristics may assign them to visual analysis either in 2D or 3D visualisations or in a combination of both (section 4.1.5). Normally, data sets showing the same variables are thought to be similar. The eight data sets employed in stage IIa all show deer tracking data of the same red deer in the Swiss Nationalpark during different times of the day and in different months of the year (section 2.5.2.3). While the variables displayed are the same the actual data values and their spatial distribution differ. Each of these eight (similar?) data sets gains in regard to selected appropriateness measures from analysis either in 2D, in 3D or in both visualisation types. Varying appropriateness of visualisation techniques based on the specific data set may also exist in other application areas where data sets may basically be thought of as being similar, for example, census data of two different years of the same country or weather data (e.g. wind or rain measurements) of different days at the same places. In some of these data sets we may expect more variation in the actual data values (e.g. weather data) in others less (e.g. census data). This research finds that if a visualisation type performs well with one data set it may perform similarly, better or worse with a similar data set depending on the actual values and/or the spatial distribution of the values. The tentative proposition about data set characteristics made in this research should be taken into account and tested further in future evaluations. Additionally, it may be useful to re-evaluate the data set characteristics of existing studies about 2D and 3D displays to complement the findings of this research. Findings from this research, re-evaluation of existing studies and specially designed future studies should ultimately lead to some guidance about what visualisation technique is most suitable for a data set. It is acknowledge, that guidance on general levels is available. For example, hierarchical data may be shown suitably as tree maps or tree graphs. Also, guidance about suitable mapping of data values to visual variables is available. For example, quantitative data should be mapped rather to symbol length than to symbol colour (Cleveland & McGill 1984, Nesbitt 2005). However, the findings of this research suggest that we may need guidance on what type of display to use based on (geo)data set characteristics such as the actual values available, for example, minimum/maximum values, the range of values, the thematic and/or spatial aggregation, as well as their density, spatial distribution and characteristics of the setting geographic data belongs to (e.g. variation in topography or land-cover).

5.5. Further research

In addition to the implications discussed above where some aspects may need further evaluation the following issues were identified during the research which should be further researched.

Visibility and readability of important background information in virtual environments: Using a traditional map as background information in the virtual environment is not perfect. The rendering quality of the virtual environment may decrease readability of textual information and symbols. Especially, the labelling of spot heights through the standard north oriented labels in the background map is not suitable and results in significantly less reported exact height values in the 3D visualisations (section 4.1.3.2). While the labels of the overlaid grid which are positioned at the boundaries of the interesting area are labelled as placemarks, and thus better visible and readable, the spot heights should be readable but nevertheless not interfere (e.g. through occlusion and overlap) with the data values. Such labels should not serve as landmarks (Harrower & Sheesley 2005). Non-interfering labels could, for example, be included as turning labels within the map symbology. Research on labelling in virtual environments is ongoing (e.g. Lehmann et al. 2011, Maas et al. 2007, Maas & Döllner 2006) but not especially for the labelling of virtual environments with data displays as explored in this research. Additional options would include interactive querying of altitude values and other information, for example, slope steepness or details about the displayed data set, or the combination of different displays. Häberling (2003) proposes a number of theses for the design of 3D maps. While some of his propositions may interfere with additional data displays they may be useful as starting points for the specific design of 3D background maps for data displays.

Interaction: Several participants and especially the data experts remarked that they would like to have more possibilities for interaction with the data sets, such as switching on/off the background frames, showing/hidding different data subsets (especially when less data variables are shown on a single data graphic), interactively querying the data set and so on. Additionally, some of the recommendations on data displays that include ideas about improvements through interaction, as discussed above, may need further research. Improvements could include, for example, changing the order of stacked bars, developing methods of varying projection and/or changing the viewpoint to avoid or minimise occlusion, interactively change the spatial and/or thematic aggregation of the data, or change the scaling of the bars or bar charts to make small bars better visible, to adapt to different zoom levels or to minimise overlap and occlusion. Generally, high interactivity is a characteristic of geovisualizations for data exploration and analysis (MacEachren & Kraak 1997). In this research, the level of interactivity increased slightly from stage I to stage III but, for example, data querying was not possible. Further evaluations would need to show how different interaction possibilities support or hinder efficient and effective visual data analysis in virtual environments. For 2D displays, Yost & North (2005, p. 1889) state that "designers of information visualization systems have the choice to present information in a single integrated view or in multiple views." Thus, instead of using increasingly complex data graphics (which are not liked by the participants, section 5.2.1.1) the use of multiple views should be considered. Participants could interactively switch between views or views are provided beside each other as discussed in the following section.

Combination of 2D and 3D displays: Some data sets in stage IIa (section 4.1.5) benefit from the visual analysis in the 2D and the 3D visualisations. Additionally, some participants ask for other displays of the same data. Options for further exploration would include the combination of two or several 2D and 3D displays concurrently on the same screen or enabling switching between these. The efficiency and effectiveness of such combined displays would need evaluation. Potentially, the 2D and/or 3D visualisations could also be combined with other data displays, for example, statistical displays (e.g. histograms or boxplots) or other map like displays (e.g. showing the not aggregated data, density or slope curvature maps) as asked for by participants. Also other types of 3D representations, for example employing orthogonal projection, could be considered for display combinations. Combined displays for data exploration should be connected by brushing (as usual in information visualisation, e.g. Roberts & Wright 2006) and potentially also allow for combined navigation. Visualisations systems of several displays and interaction functionality should also usefully allow for workspace and workflow support, for example, by maintaining a history of actions, allowing annotations and supporting different strands of data exploration. These are not new research areas. In 2D information visualisation, combined and interactively connected 2D displays are often used for exploratory data analysis tasks (e.g. Godinho, Meiguins, Meiguins, Casseb do Carmo, Garcia, Almeida & Lourenco 2007, Andrienko & Andrienko 2006, Roberts 2008). Other researchers (Kreuseler 2000, Dykes, Moore & Wood 1999, Hetherington, Farrimond & Clynch 2007, Chang et al. 2007, Beard, Hay, Nicoll & Edge 2005) have integrated 2D and 3D views in combined visualisation displays for data and information exploration and communication. Several studies (section 1.2.2) have also shown that a combination of 2D and 3D displays can be useful for gaining insight into the data set and/or for the easier navigation especially of the 3D visualisation, for example with a 2D overview window. Ware & Plumlee (2005) explicitly recommend providing extra windows and/or views in 3D geovisualization displays when more complex objects or patterns need comparison. Visualising the same data differently has the potential to enrich the data analysis (e.g. by re-expression, DiBiase et al. 1992). Additionally, participants can use the display(s) they like more. The effect of personal preference in this research was quantitative rather than qualitative (section 5.2.1.4). Further research on the influence of personal preferences may be helpful. Other researchers (Hegarty, Smallman & Stull 2008) have shown that participants often prefer displays which do not help their performance. This could be tested for by a quick user validation process as suggested in section 5.4. Additionally, combined displays would leave participants who defended 2D displays strongly and would not like to see them (completely) replaced at ease (section 3.6.4). First experiments with the combination and connection of data displays in 3D virtual environments as tested in this research with statistical 2D data displays are shown in Bleisch & Nebiker (2008) but are not formally evaluated yet. 3D visualisations should be viewed not as a replacement but rather a supplement to traditional and future 2D representations. Each display type may play out its strengths when suitable and combined displays may even allow a combination of strengths of single display types. However, future evaluations will need to establish if the costs of additional displays, and thus potentially more complicated navigation and/or interaction, does not outweigh the potential benefits in efficiency and/or effectiveness of gaining insights into data sets.

Different data graphics: Further research may also explore the suitability of different data graphic types, for example, 3D bars, circles, cones, etc., for the visual analysis of complex data sets and tasks in virtual environments (section 4.1.1). First evaluations of 3D bars and 2D circles within virtual environments have shown that 3D bars are similarly effective for simple tasks as 2D bars on billboards especially when using reference frames (Bleisch accepted). However, as this research has shown, reference frames crowd the display when data sets are more dense. As a reference frame of a 3D bar consists of many more black lines than a reference frame of a 2D bar the problem of too crowded displays may get worse when employing 3D bars with reference frames to display varying quantitative data values in virtual environments. But for simpler displays 3D bars may be appropriate as they may better integrate in the else three-dimensional landscape than 2D bars on billboards do.

5.6. Outlook

This research effectively evaluated the visual analysis of quantitative data displays in relation to the landscape the data sets belong to. While data analysis in relation to altitude and landform is a possibility it has rarely been done so far. But most participants and especially the data experts are open to the scope offered by 3D visualisation techniques. Already, and certainly in the future, sensor networks and other data collection means produce an ever increasing amount and density of data often related to the three-dimensional position where the data was collected. Thus, the potential for three-dimensional visual analysis of data sets will increase. Potentially, not only data in relation to landscape and landscape features will need analysis but also data collected in and around other objects, for example, measurements of wind speed and temperature at specific positions on large buildings, bridges or other constructions. ”Smart structures” (Economist 2010), structures ’peppered’ with sensors, can provide continuous monitoring of the structure and its environment. Additionally, (harmful) substances, for example ozone or volcanic ash, are measured midair by drones or other substances in the oceans and other waters. All this collected data may benefit from visual analysis in virtual environments as the measurements are displayed, for example as data graphics, at the actual three-dimensional positions where they are collected in. They can be visually analysed three-dimensionally within the data set, in relation to other data sets displayed and also in relation to objects, the landscape or submarine or even subterranean features. A few of these ideas were informally discussed with the case study data experts who considered them very interesting. Somehow, visual data analysis in three dimensions seems a natural extension to the very successful concept of visually analysing (geo)data in two dimensions, for example on maps.

For data along 3D structures the display of a single variable could be rendered onto the surface of the structure (Lorenz & Döllner 2010). But this is less appropriate for multivariate data, data collected midair or data without a specific relation to the surface. A further challenge will be the extension of three-dimensional displays with suitable representations of time and data changes over time. Current research on the visualisation of time (e.g. Kraak 2008, Li & Kraak 2008) may provide the basis for such extensions.

Some years ago Funamizu (2008) presented a vision of a mobile device for internet search and result visualisation (figure 5.1). Such or similar devices for augmented reality applications may also gain from suitable displays of data and information in three- or four-dimensional space and, thus, data and information which is correctly located in space (and time) and readily visible and interpretable.

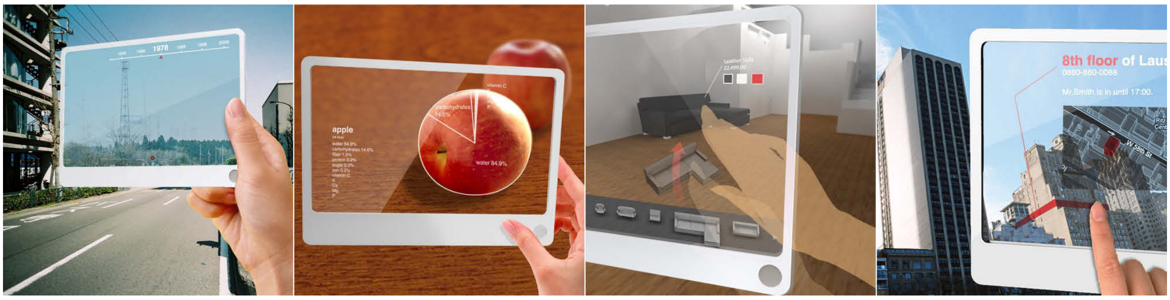


Figure 5.1.: Visionary mobile device augmenting reality with data and information (Funamizu 2008)

References

The references are ordered alphabetically by the surname of the first author. The provided URLs were working on April 2nd, 2011.

Akenine-Möller, T. & Haines, E. (2002), *Real-Time Rendering*, 2nd edn, A K Peters, Natick.

Amar, R., Eagan, J. & Stasko, J. (2005), Low-Level Components of Analytic Activity in Information Visualization, *in* IEEE Symposium on Information Visualization (INFOVIS'05)', IEEE, Minneapolis, pp. 111–117.

Ambroziak, B. M. & Ambroziak, J. R. (1999), *Infinite Perspectives: Two Thousand Years of Three-Dimensional Mapmaking*, Princeton Architectural Press, New York.

Andrienko, N. & Andrienko, G. (2006), *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*, Springer, Berlin.

Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J. & Kretchmer, U. (2002), Testing the Usability of Interactive Maps in CommonGIS', *Cartography and Geographic Information Science* **29**(4), 325–342.

Apache (2010), SVG Rasterizer', http://xmlgraphics.apache.org/batik/tools/rasterizer.html.

ATLAS.ti (2010), ATLAS.ti - The Qualitative Data Analysis Software', http://www.atlasti.com/.

Autodesk (2009), Autodesk LandXplorer 2009', http://www.3dgeo.de/citygml.aspx.

Bair, A. & House, D. (2007), A Grid with a View: Optimal Texturing for Perception of Layered Surface Shape', *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1656–1663.

Baird, J. C. & Noma, E. (1978), *Fundamentals of Scaling and Psychophysics*, Wiley Series in Behavior, Wiley, New York.

Ball, R. & North, C. (2008), The effects of peripheral vision and physical navigation on large scale visualization, *in* ACM International Conference Proceeding Series', Vol. 322, pp. 9–16.

Bartoschek, T. & Schönig, J. (2008), Trends und Potenziale von virtuellen Globen in Schule, Lehramtsausbildung und Wissenschaft', *GIS Science* (4), 28–31.

Basdogan, C., Sedef, M., Harders, M. & Wesarg, S. (2007), VR-Based Simulators for Training in Minimally Invasive Surgery', *IEEE Computer Graphics and Applications* **27**(2), 54–66.

Batty, M. & Smith, A. (2002), Virtuality and cities: Definitions, geographies, designs, *in* P. Fisher & D. J. Unwin, eds, Virtual Reality in Geography', Vol. Chapter 19, Taylor & Francis, London, pp. 270–291.

Bazeley, P. (2004), Issues in Mixing Qualitative and Quantitative Approaches to Research, *in* R. Buber, J. Gadner & L. Richards, eds, Applying qualitative methods to marketing management research', Palgrave Macmillan, UK, pp. 141–156.

Beard, D. J., Hay, R. J., Nicoll, M. G. & Edge, D. O. (2005), 3D Web Mapping - 3D Geoscience Information Online, *in* SSC 2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute', Melbourne.

Berlin (2010), 3D-Stadtmodell Berlin', http://www.virtual-berlin.de.

Blackboard (2010), Blackboard International', http://www.blackboard.com/International/EMEA.aspx?lang=en-us.

Bleisch, S. (2007), Towards an efficient visual combination of abstract quantitative data representations with virtual environments, *in* MPhil-PhD Upgrade Seminar', City University London, London.

Bleisch, S. (2008), Towards an appropriate visual combination of abstract quantitative data representations with virtual environments, *in* Doktorandenkolloquium FHNW & DLR', Muttenz.

Bleisch, S. (2009), Data visualization proposal for mountainous environments, *in* 3. Anwendertreffen GIS in Nationalen Naturlandschaften, 2. Workshop GIS within the network of alpine protected areas', Schweizerischer Nationalpark, Zernez, Switzerland.

Bleisch, S. (2010*a*), Informationsvisualisierung in 3D-Geoinformationsumgebungen, *in* geosuisse Winterveranstaltung', Zürich.

Bleisch, S. (2010*b*), Visualisierung quantitativer Daten in virtuellen 3D-Landschaften , *in* Kartographisches Kolloquium', Hochschule Karlsruhe - Technik und Wirtschaft, Fakultät für Geomatik Deutsche Gesellschaft für Kartographie, Sektion Karlsruhe, Karlsruhe.

Bleisch, S. (2010*c*), Visualisierung quantitativer Daten in virtuellen 3D-Landschaften , *in* FHNW HABG BBL', Muttenz.

Bleisch, S. (accepted), Towards appropriate representations of quantitative data in virtual environments', *Cartographica* **xx**(xx), xx.

Bleisch, S., Burkhard, J. & Nebiker, S. (2009), Efficient Integration of Data Graphics into Virtual 3D Environments, *in* 24th International Cartography Conference', Santiago de Chile.

Bleisch, S. & Dykes, J. (2008), Using Web-Based 3-D Visualization for Planning Hikes Virtually: An Evaluation, *in* N. J. Mount, G. L. Harvey, P. Aplin & G. Priestnall, eds, Representing, Modeling and Visualizing the Natural Environment: Innovations in GIS 13', CRC Press, Florida, pp. 353–365.

Bleisch, S., Dykes, J. & Nebiker, S. (2006), How are we to Represent Abstract Information in 'Realistic' 3D Environments?, *in* GIScience 2006', Münster, Germany.

Bleisch, S., Dykes, J. & Nebiker, S. (2007), Evaluating the effectiveness of representing numeric information through abstract graphics in desktop 3D environments, *in* ICA Commission on Visualization and Virtual Environments Workshop', Helsinki.

Bleisch, S., Dykes, J. & Nebiker, S. (2008), Evaluating the Effectiveness of Representing Numeric Information Through Abstract Graphics in 3D Desktop Virtual Environments', *The Cartographic Journal* **45**(3), 216–226.

Bleisch, S., Dykes, J. & Nebiker, S. (2009*a*), A meta-framework of methodological approaches exemplified by 3D geovisualization research, *in* Workshop Human Aspects of Visualization, INTERACT 2009', Uppsala, Schweden.

Bleisch, S., Dykes, J. & Nebiker, S. (2009*b*), Building bridges between methodological approaches: a meta-framework linking experiments and applied studies in 3D geovisualization research, *in* GIS Research UK 17th Annual Conference', Durham, UK.

Bleisch, S., Dykes, J. & Nebiker, S. (2010*a*), A mixed methods research approach for 3D geovisualization evaluation, *in* Workshop on Methods and Techniques of Use, User and Usability Research', UCL, London.

Bleisch, S., Dykes, J. & Nebiker, S. (2010*b*), Forest or slope? Comparing 2d and 3d visualisations through the wording of task answers, *in* GIScience 2010, Sixth international conference on Geographic Information Science', Zürich.

Bleisch, S. & Nebiker, S. (2008), Connected 2D and 3D visualizations for the interactive exploration of spatial information, *in* XXI ISPRS Congress', Beijing, China.

Bodum, L. (2005), Modelling Virtual Environment for Geovisualization: A Focus on Representation, *in* J. Dykes, A. M. MacEachren & M.-J. Kraak, eds, Exploring Geovisualization', Vol. Chapter 19, Elsevier on behalf of the International Cartographic Association ICA, Amsterdam, pp. 389–402.

Boren, T. M. & Ramey, J. (2000), Thinking aloud: reconciling theory and practice', *IEEE Transactions on Professional Communication* **43**(3), 261–278.

Botts, M., Percivall, G., Reed, C. & Davidson, J. (2006), OGC Sensor Web Enablement: Overview and High Level Architecture', `http://www.opengeospatial.org/projects/groups/sensorweb`.

Bowman, D. A., North, C., Chen, J., Polys, N. F., Pyla, P. S. & Yilmaz, U. (2003), Information-rich virtual environments: theory, tools and research agenda, *in* A. SIGGRAPH, ed., Virtual Reality Software and Technology, Prodeedings of the ACM symposium on Virtual reality software and technology', ACM Press, Osaka, Japan, pp. 81–90.

Brath, R. (1997), Concept Demonstration Metrics for Effective Information Visualization, *in* J. Dill & N. Gershon, eds, IEEE Symposium on Information Visualization', IEEE, Phoenix, pp. 108–126.

Brooks, S. & Whalley, J. L. (2008), Multilayer hybrid visualizations to support 3D GIS', *Computers, Environment and Urban Systems* **32**(4), 278–292.

Burkhard, J. (2008), XML-basierte 3D-Datenaufbereitung, MSc, FHNW Fachhochschule Nordwestschweiz.

Butler, D. (2006), The web-wide world', *Nature* **439**(16), 776–778.

Campell, S. & Filli, F. (2006), Habitat Selection and Habitat Use of Female Chamois Rupicapra rupicapra in Winter'.

Cartwright, W., Crampton, J., Gartner, G., Miller, S., Mitchell, K., Siekierska, E. & Wood, J. (2001), Geospatial Information Visualization User Interface Issues', *Cartography and Geographic Information Science* **28**(1).

Cawthon, N. & Vande Moere, A. (2007), The Effect of Aesthetic on the Usability of Data Visualization, *in* E. Banissi, R. A. Burkhard, G. Grinstein, U. Cvek, M. Trutschl, L. Stuart, T. G. Wyeld, G. Andrienko, J. Dykes, M. Jern, D. Groth & A. Ursyn, eds, 11th International Conference Information Visualization (IV'07)', IEEE, Zürich, pp. 637–645.

Cerny, J. W. & Wilson, J. (1976), The effect of orientation on the recognition of simple maps', *The Canadian Cartographer* **13**(2), 132–138.

Chang, R., Wessel, G., Kosara, R., Sauda, E. & Ribarsky, W. (2007), Legible Cities: Focus-Dependent Multi-Resolution Visualization of Urban Relationships', *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1169–1175.

Chen, C. & Yu, Y. (2000), Empirical studies of information visualization: a meta-analysis', *Int. J. Human-Computer Studies* **53**, 851–866.

Cleveland, W. S. (1994), *The Elements of Graphing Data*, 2nd edn, Hobart Press, New Jersey.

Cleveland, W. S. & McGill, R. (1984), Graphical Percpetion: Theory, Experimentation, and Application to the Development of Graphical Methods', *Journal of the American Statistical Association* **79**(387), 531–554.

Cockburn, A. (2004), Revisiting 2D vs 3D Implications on Spatial Memory, *in* Proceedings of the fifth conference on Australasian user interface', Vol. 28, ACM, Dunedin, New Zealand, pp. 25–31.

Cockburn, A. & McKenzie, B. (2001), 3D or not 3D? Evaluating the Effect of the Third Dimension in a Document Management System, *in* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', ACM, Seattle.

Cockburn, A. & McKenzie, B. (2004), Evaluating spatial memory in two and three dimensions', *Int. J. Human-Computer Studies* **61**(2004), 359–373.

Collier, P., Forrest, D. & Pearson, A. (2003), The Representation of Topographic Information on Maps: The Depiction of Relief', *The Cartographic Journal* **40**(1), 17–26.

Cowie, A. C., ed. (1989), *Oxford Advanced Learner's Dictionary of current English*, 4th edn, University Press, Oxford.

Craglia, M., Goodchild, M. F., Annoni, A., Camara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S. & Parsons, E. (2008), Next-Generation Digital Earth', *International Journal of Spatial Data Infrastructures Research* **3**, 146–167.

Creswell, J. W., Plano Clark, V. L. & Garrett, A. L. (2008), Methodological Issues in Conducting Mixed Methods Research Designs, *in* M. M. Bergman, ed., Advances in Mixed Methods Research - Theories and Applications', SAGE, London, pp. 66–84.

Dawood, N. & Sikka, S. (2007), Measuring the Effectiveness of 4D Planning as a Valuable Communication Tool, *in* 7th International Conference on Construction Applications of Virtual Reality', pp. 198–205.

de Boer, A. (2008), Vertical exaggeration of prestigious buildings in historical city maps', `http://arnouddeboer.geomultimedia.nl/node/14`.

DiBiase, D., MacEachren, A. M., Krygier, J. B. & Reeves, C. (1992), Animation and the Role of Map Design in Scientific Visualization', *Cartography and Geographic Information Systems* **19**(4), 201–214, 265–266.

Döllner, J. & Walther, M. (2003), Real-Time Expressive Rendering of City Models, *in* Proceedings of the Seventh International Conference on Information Visualization (IV'03)', IEEE, London.

Dörries, M. (2005), Krakatau 1883: Die Welt als Labor und Erfahrungsraum, *in* I. Schröder & S. Höhler, eds, Welt-Raeume: Geschichte, Geographie und Globalisierung seit 1900', Campus Verlag, pp. 51–73.

Dykes, J., Moore, K. & Wood, J. (1999), Virtual environments for student fieldwork using networked components', *International Journal of Geographical Information Science* **13**(4), 397–416.

Economist (2010), Superstructures', `http://www.economist.com`.

Edsall, R. M. & Deitrick, S. (2009), Case studies demonstrating the utility of uncoventional designs for geographic problem-solving, *in* ICC International Cartographic Conference', Santiago de Chile, Chile.

Elmes, D. G., Kantowitz, B. H. & Roediger, H. L. (2006), *Research Methods in Psychology*, 8th edn, Thomson Wadsworth, Belmont.

Elmqvist, N. & Tsigas, P. (2008), A Taxonomy of 3D Occlusion Management for Visualization', *IEEE Transactions on Visualization and Computer Graphics* **14**(5), 1095–1109.

Elmqvist, N. & Tudoreanu, M. E. (2007), Occlusion Management in Immersive and Desktop 3D Virtual Environments: Theory and Evaluation', *The International Journal of Virtual Reality* **6**(2), 21–32.

ERDAS (2010), ERDAS TITAN Client', http://www.erdas.com/products/ERDASTITANClient/Details.aspx.

ESRI (2003), Introducing ArcGlobe - An ArcGIS 3D Analyst Application', http://www.esri.com/news/arcnews/summer03articles/introducing-arcglobe.html.

Fabrikant, S. I. & Montello, D. R. (2008), The effect of instruction on distance and similarity judgements in information spatializations', *International Journal of Geographical Information Science* **22**(4), 463–478.

Fabrikant, S. I., Montello, D. R. & Mark, D. M. (2006), The Distance-Similarity Metaphor in Region-Display Spatialization', *IEEE Computer Graphics and Applications* **26**(4), 34–44.

Fairbairn, D., Andrienko, G., Andrienko, N., Buziek, G. & Dykes, J. (2001), Representation and its relationship with cartographic visualization: a research agenda', *Cartography and Geographic Information Science* **28**(1).

Few, S. (2009), *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, Analytics Press.

Filli, F. & Suter, W., eds (2006), *Huftierforschung im Schweizerischen Nationalpark / Ungulate Research in the Swiss National Park*, Vol. 93 of *Nationalpark-Forschung in der Schweiz*, Forschungskommission des Schweizerischen Nationalparks (SCNAT Kommission), Zernez.

Flannery, J. J. (1971), The relative effectiveness of some common graduated point symbols in the presentation of quantitative data', *The Canadian Cartographer* **8**(2), 96–109.

Foddy, W. (1994), *Constructing questions for interviews and questionnaires: theory and practice in social research*, Cambridge University Press.

Forsberg, A., Chen, J. & Laidlaw, D. (2009), Comparing 3D Vector Field Visualization Methods: A User Study', *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 1219–1226.

Forsberg, A., Prabhat, Haley, G., Bragdon, A., Levy, J., Fassett, C. I., Shean, D., Head, J. W., Milkovich, S. & Duchaineau, M. A. (2006), Adviser: Immersive Field Work for Planetary Geoscientists', *IEEE Computer Graphics and Applications* **26**(4), 46–54.

FOXNews.com (2010), Ancient Street Uncovered in Jerusalem', http://www.foxnews.com/scitech/2010/02/10/ancient-street-uncovered-jerusalem/.

Fuhrmann, S., Ahonen-Rainio, P., Edsall, R. M., Fabrikant, S. I., Koua, E. L., Tobon, C., Ware, C. & Wilson, S. (2005), Making Useful and Useable Geovisualization: Design and Evaluation Issues, *in* J. Dykes, A. M. MacEachren & M.-J. Kraak, eds, Exploring Geovisualization', Vol. Chapter 28, Elsevier on behalf of the International Cartographic Association ICA, Amsterdam, pp. 553–566.

Fuller, R. B. (1969), *Operating Manual for Spaceship Earth*, Southern Illinois University Press, Carbondale.

Funamizu, M. (2008), Future of Internet Search: Mobile version', http://petitinvention.wordpress.com/2008/02/10/future-of-internet-search-mobile-version/.

G2007 (2010), Deformationsmessungen "Schwanderbärgli", Schwanden bei Brienz', http://web.fhnw.ch/habg/projekte/brienz/.

Gabbard, J. L., Hix, D. & Swan, J. E. (1999), User-Centered Design and Evaluation of Virtual Environments', *IEEE Computer Graphics and Applications* **Nov/Dez**, 51–59.

Gahegan, M. (1999), Four barriers to the development of effective exploratory visualisation tools for the geosciences', *Int. J. Geographical Information Science* **13**(4), 289–309.

Gahegan, M., Wachowicz, M., Harrower, M. & Rhyne, T.-M. (2001), The Integration of Geographic Visualization with Knowledge Discovery in Databases and Geocomputation', *Cartography and Geographic Information Science* **28**(1), 29–44.

Geng, L. & Hamilton, H. J. (2006), Interestingness Measures for Data Mining: A Survey', *ACM Computing Surveys* **38**(3), 1–32.

GeoICT (2007), Sensor Web @ GeoICT', http://sensorweb.geoict.net/.

Gerring, J. (2007), *Case Study Research - Principles and Practices*, University Press, Cambridge.

Gerring, J. & McDermott, R. (2007), An Experimental Template for Case Study Research', *American Journal of Political Science* **51**(3), 688–701.

Godinho, P. I. A., Meiguins, B. S., Meiguins, A. S. G., Casseb do Carmo, R. M., Garcia, M. d. B., Almeida, L. H. & Lourenco, R. (2007), PRISMA - A Multidimensional Information Visualization Tool Using Multiple Coordinated Views, *in* 11th International Conference Information Visualization (IV'07)', IEEE.

Google (2010), Google Earth', http://earth.google.com/.

Gore, A. (1998), The Digital Earth: Understanding our planet in the 21st Century', http://portal.opengeospatial.org/files/?artifact_id=6210.

Greenberg, S. & Buxton, B. (2008), Usability Evaluation Considered Harmful (Some of the Time), *in* 2008 ACM Conference on Human Factors in Computing Systems, CHI'08', Florence, Italy.

Greene, J. C., Caracelli, V. J. & Graham, W. F. (1989), Toward a Conceptual Framework for Mixed-Method Evaluation Design', *Educational Evaluation and Policy Analysis* **11**(3), 255–274.

Gross, N. (1999), The earth will don an electronic skin', http://www.businessweek.com/1999/99_35/b3644024.htm.

Häberling, C. (2003), Topografische 3D-Karten - Thesen für kartographische Gestaltungsgrundsätze, PhD, ETH Zürich.

Hägerstrand, T. (1970), What about people in regional science?', *Papers of the Regional Science Association* **24**(1), 6–21.

Harrower, M. & Sheesley, B. (2005), Moving Beyond Novelty: Creating Effective 3D Fly-over Maps, *in* ICC International Cartographic Conference', ICA, A Coruna.

Harvey, F. (2009), More than Names - Digital Earth and/or Virtual Globes?', *International Journal of Spatial Data Infrastructures Research* **4**, 111–116.

Haskell, I. D. & Wickens, C. D. (1993), Two- and Three-Dimensional Displays for Aviation: A Theoretical and Empirical Comparison', *The International Journal of Aviation Psychology* **3**(2), 87–109.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K. & Subbiah, I. (2002), Development of a self-report measure of environmental spatial ability', *Intelligence* **30**, 425–447.

Hegarty, M., Smallman, H. S. & Stull, A. T. (2008), Decoupling of Intuitions and Performance in the Use of Complex Visual Displays, *in* CogSci 2008 30th Annual Meeting of the Cognitive Science Society', Washington, DC, USA, pp. 881–886.

Hetherington, R., Farrimond, B. & Clynch, P. (2007), Interactive Web Visualisation of Proposals for Site Developments, *in* E. Banissi, R. A. Burkhard, G. Grinstein, U. Cvek, M. Trutschl, L. Stuart, T. G. Wyeld, G. Andrienko, J. Dykes, M. Jern, D. Groth & A. Ursyn, eds, 11th International Conference Information Visualization, IV2007', IEEE Computer Society, Zürich, Switzerland, pp. 613–622.

Hiebel, G., Hanke, K. & Hayek, I. (2010), Räumliches Informationssystem auf Grundlage einer ontologiebasierten Datenstruktur für multidisziplinäre Forschung, *in* DGPF Tagungsband 19/2010 - Dreiländertagung OVG, DGPF und SGPF', Wien, pp. 685–694.

Hollands, J. G. & Ivanovic, N. (2002), Task Switching with 2D and 3D Displays of Geographic Terrain: The Role of Visual Momentum, *in* RTO IST Workshop on "Massive Military Data Fusion and Visualisation: Users Talk with Developers"', Halden, Norway.

Hübsch, C. (2008), XQuest: Fragebogenerstellung und -auswertung', http://chu.in-chemnitz.de/programmieren/xslt/.

Hutchins, E. (1995), *Cognition in the Wild*, MIT Press, Cambridge.

Inselberg, A. (2008), Parallel Coordinates: Visualization, Exploration and Classification of High-Dimensional Data, *in* C.-h. Chen, W. K. Härdle & A. Unwin, eds, Handbook of Data Visualization', Springer Handbooks of Computational Statistics, Springer, pp. 643–680.

Irani, P. & Ware, C. (2003), Diagramming Information Structures Using 3D Perceptual Primitives', *ACM Transactions on Computer-Human Interaction* **10**(1), 1–19.

Jankun-Kelly, T. (2006), Is There Science in Visualization?, *in* IEEE Visualization 2006', Baltimore, USA.

Jansen-Osmann, P. & Berendt, B. (2002), Investigating Distance Knowledge Using Virtual Environments', *Environment and Behavior* **34**(2), 178–193.

Jianu, R., Demiralp, C. & Laidlaw, D. H. (2009), Exploring 3D DTI Fiber Tracts with Linked 2D Representations', *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 1449–1456.

Jobst, M. & Germanchis, T. (2007), The Employment of 3D in Cartography - An Overview, *in* W. Cartwright, M. P. Peterson & G. Gartner, eds, Multimedia Cartography', 2nd edn, Springer, Berlin, pp. 217–228.

Johns, C. (2003), Spatial learning: cognitive mapping in abstract virtual environments, *in* 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa', ACM, Cape Town, South Africa, pp. 7–16.

Johnson, C. (2004), Top Scientific Visualization Research Problems', *IEEE Computer Graphics and Applications* **24**(4), 13–17.

Johnson, C., Moorhead, R., Munzner, T., Pfister, H., Rheingans, P. & Yoo, T. S., eds (2006), *NIH/NSF Visualization Research Challenges*, IEEE Press.

Johnson, R. B. & Onwuegbuzie, A. J. (2004), Mixed Methods Research: A Research Paradigm Whose Time Has Come’, *Educational Researcher* **33**(7), 14–26.

Johnson, R. B., Onwuegbuzie, A. J. & Turner, L. A. (2007), Toward a Definition of Mixed Methods Research’, *Journal of Mixed Methods Research* **1**(2), 112–133.

Jones, R., McCaffrey, K., Clegg, P., Wilson, R., Holliman, N., Holdsworth, R., Imber, J. & Waggott, S. (2007), Integration of regional to outcrop digital data: 3D visualisation of multi-scale geological models’, *Computers & Geosciences* **33**.

JPEG (2010), The JPEG committee home page’, http://www.jpeg.org/.

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P. & Montello, D. R. (2008), Spatial Reasoning With External Visualizations: What Matters Is What You See, Not Whether You Interact’, *Cognitive Science* **32**(2008), 1099–1132.

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

Keller, P. R. & Keller, M. M. (1993), *Visual Cues: Practical Data Visualization*, IEEE Computer Society Press, Los Alamitos.

Kettunen, P., Sarjakoski, T., Sarjakoski, L. T. & Oksanen, J. (2010), A Cartographic 3D View in Oblique Parallel Projection’, http://www.vector1media.com/article/features/13536-a-cartographic-3d-view-in-oblique-parallel-projection.html.

Kimerling, A. J., Buckley, A. R., Muehrcke, P. C. & Muehrcke, J. O. (2009), *Map Use: Reading and Analysis*, 6th edn, ESRI Press, Redlands.

Koua, E. L., MacEachren, A. M. & Kraak, M.-J. (2006), Evaluating the usability of visualization methods in an exploratory geovisualization environment’, *International Journal of Geographical Information Science* **20**(4), 425–448.

Kraak, M.-J. (1988), *Computer-assisted Cartographical Three-dimensional Imaging Techniques*, Delft University Press, Delft.

Kraak, M.-J. (2006), Beyond Geovisualization’, *IEEE Computer Graphics and Applications* **July**/**August**, 6–9.

Kraak, M.-J. (2008), Geovisualization and Time - New Opportunities for the Space-Time Cube, *in* M. Dodge, M. McDerby & M. Turner, eds, Geographic Visualization: Concepts, Tools and Applications’, Wiley, Chichester, pp. 293–306.

Kreuseler, M. (2000), Visualization of geographically related multidimensional data in virtual 3D scenes’, *Computers & Geosciences* **26**(1), 101–108.

Kumke, H. (2009), Darstellung thermaler numerischer Informationen auf dreidimensionalen Gebäudefassaden, *in* GeoViz2009’, Hamburg.

Lange, E. (1999), The Degree of Realism of GIS-Based Virtual Landscapes: Implications for Spatial Planning, *in* D. Fritsch & R. Spiller, eds, Photogrammetric Week 99’, Wichmann Verlag, Heidelberg, pp. 367–374.

Lehmann, C., Trümper, J. & Döllner, J. (2011), Interactive Areal Annotations for 3D Treemaps of Large-Scale Software Systems, *in* GeoViz 2011’, Hamburg.

Leica (2005), Leica Virtual Explorer - Comprehensive 3D Visualization Suite’, http://gi.leica-geosystems.com/documentcenter/lve/VirtualExplorer_Brochure.pdf.

Lewins, A. & Silver, C. (2007), *Using Software in Qualitative Research: A Step-by-Step Guide*, Sage Publications Ltd., London.

Li, X. & Kraak, M.-J. (2008), The time wave: a new method of visual exploration of geo-data in time-space’, *The Cartographic Journal* **45**(3), 193–200.

Lloyd, D. (2009), Evaluating human-centered approaches for geovisualization, PhD, City University London.

Lorenz, A., Thierbach, C., Kolbe, T. H. & Baur, N. (2010), Untersuchung der Effizienz und Akzeptanz von 2D- und 3D-Kartenvarianten für die Innenraumnavigation, *in* DGPF Tagungsband 19/2010 - Dreiländertagung OVG, DGPF und SGPF’, Wien, pp. 342–355.

Lorenz, H. & Döllner, J. (2010), 3D feature surface properties and their application in geovisualization’, *Computers, Environment and Urban Systems* **34**(2010), 476–483.

Maas, S. & Döllner, J. (2006), Ein Konzept zur dynamischen Annotation virtueller 3D-Stadtmodelle, *in* GEOVIS', Vol. Band 10: Aktuelle Entwicklungen in Geoinformation und Visualisierung of *Kartographische Schriften*, Deutsche Gesellschaft für Kartographie, Potsdam, pp. 19–26.

Maas, S., Jobst, M. & Döllner, J. (2007), Use of depth cues for the annotation of 3D geo-virtual environments, *in* International Cartographic Conference ICC 2007', Moscow, Russia.

MacEachren, A. M. (1994), Visualization in Modern Cartography: Setting the Agenda, *in* A. M. MacEachren & D. R. F. Taylor, eds, Visualization in Modern Cartography', Vol. 2, Pergamon, Oxford, pp. 1–12.

MacEachren, A. M. (1995), *How Maps Work - Representation, Visualization, and Design*, The Guilford Press, New York.

MacEachren, A. M., Edsall, R., Haug, D., Baxter, R., Otto, G., Masters, R., Fuhrmann, S. & Qian, L. (1999*a*), Virtual environments for geographic visualization: Potential and challenges, *in* Proceedings of the ACM Workshop on New Paradigms in Information Visualization and Manipulation', KS, Kansas City, pp. 35–40.

MacEachren, A. M., Edsall, R., Haug, D., Baxter, R., Otto, G., Masters, R., Fuhrmann, S. & Qian, L. (1999*b*), Virtual environments for geographic visualization: potential and challenges, *in* 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM internation conference on Information and knowledge management', ACM, Kansas City, pp. 35–40.

MacEachren, A. M. & Kraak, M.-J. (1997), Exploratory Cartographic Visualization: Advancing the Agenda', *Computers & Geosciences* **23**(4), 335–343.

MacEachren, A. M. & Kraak, M.-J. (2001), Research Challenges in Geovisualization', *Cartography and Geographic Information Science* **28**(1), 3–12.

MacEachren, A. M., Kraak, M.-J. & Verbree, E. (1999), Cartographic issues in the design and application of geospatial virtual environments, *in* 19th International Cartographic Conference', Ottawa, Canada.

Mackinlay, J. (1986), Automating the design of graphical presentations of relational information', *ACM Transactions on Graphics* **5**(2), 110–141.

Marsh, S. L. (2007), Using and Evaluating HCI Techniques in Geovisualization: Applying Standard and Adapted Methods in Research and Education, PhD, City University.

MAXON (2011), Cinema 4D', `http://www.maxon.net`.

mediamachines (2010), Flux Player and Flux Studio', `http://mediamachines.wordpress.com/flux-player-and-flux-studio/`.

Medyckyj-Scott, D. (1994), Visualization and Human-computer Interaction in GIS, *in* H. M. Hearnshaw & D. J. Unwin, eds, Visualizations in Geographical Information Systems', Wiley, Chichester, pp. 200–211.

Mejdal, S., McCauley, M. E. & Beringer, D. B. (2001), Human Factors Design Guidelines for Multifunction Displays, Report DOT/FAA/AM-01/17, U.S. Department of Transportation, Federal Aviation Administration.

Meng, L. (2003), Missing Theories and Methods in Digital Cartography, *in* 21st International Cartographic Conference', Durban.

Meyer, D. L. & Filli, F. (2006), Summer and Winter Ranges of Red Deer Hinds Cervus elaphus in the Swiss National Park'.

Microsoft (2010*a*), Bing Maps 3D', `http://www.bing.com/maps/`.

Microsoft (2010*b*), Microsoft Excel', `http://office.microsoft.com/de-ch/excel/`.

Miles, M. B. & Huberman, A. M. (1994), *Qualitative Data Analysis*, 2nd edn, Sage Publications, Thousand Oaks.

Mine, M. (2003), Towards Virtual Reality for the Masses: 10 Years of Research at Disney's VR Studio, *in* J. Deisinger & A. Kunz, eds, International Immersive Projection Technologies Workshop & Eurographics Workshop on Virtual Environments', The Eurographics Association, pp. 11–17.

Mitas, L., Brown, W. M. & Mitasova, H. (1997), Role of dynamic cartography in simulations of landscape processes based on multi-variate fields', *Computers and Geosciences* **23**(4), 437–446.

Montello, D. R. (2002), Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches', *Cartography and Geographic Information Science* **29**(3), 283–304.

Morrison, K. W. & Purves, R. S. (2002), Customizable landscape visualizations. Implementation, application and testing of a web-based tool', *Computers, Environment and Urban Systems* **26**(2002), 163–183.

Morse, E., Lewis, M. & Olsen, K. A. (2000), Evaluating visualizations: using a taxonomic guide', *Int. J. Human-Computer Studies* **53**, 637–662.

Morville, P. (2005), *Ambient Findability*, O'Reilly, Sebastopol, CA.

NASA (2008), NASA World Wind', http://worldwind.arc.nasa.gov/.

NASA (2010), NASA World Wind Java Demo Applications and Applets', http://worldwind.arc.nasa.gov/java/demos/.

Nebiker, S., Bleisch, S. & Christen, M. (2010), Rich point clouds in virtual globes - A new paradigm in city modeling?', *Computers, Environment and Urban Systems* **34**(6), 508–517.

Nebiker, S. & Christen, M. (2010), i3D Virtual Globe Technologie ', http://www.fhnw.ch/habg/ivgi/forschung/i3d/virtueller-globus-i3d.

Nebiker, S., Christen, M., Eugster, H., Flückiger, K. & Stierli, C. (2007), Integrating mobile geo sensors into collaborative virtual globes - design and implementation issues, *in* 5th International Symposium on Mobile Mapping Technology MMT'07', Padua.

Nebiker, S., Gülch, E. & Bleisch, S. (2010), Virtual Globes, Technical report, EuroSDR.

Nekrasovski, D., Bodnar, A., McGrenere, J., Guimbretiere, F. & Munzner, T. (2006), An Evaluation of Pan&Zoom and Rubber Sheet Navigation with and without an Overview, *in* CHI 2006 Proceedings', Montreal, Quebec, Canada.

Nesbitt, K. V. (2005), Using Guidelines to assist in the Visualisation Design Process, *in* Asia Pacific Symposium on Information Visualisation (APVIS 2005)', Sydney, Australia.

Neuendorf, K. A. (2002), *The Content Analysis Guidebook*, Sage Publications, Inc., Thousand Oaks.

NGA (2010), World Geodetic System', https://www1.nga.mil/ProductsServices/GeodesyGeophysics/WorldGeodeticSystem/Pages/default.aspx.

Nielsen, A. (2007), A Qualification of 3D Geovisualisation, PhD, Aalborg University.

Nielson, J., Clemmensen, T. & Yssing, C. (2002), Getting access to what goes on in people's heads?: reflections on the think-aloud technique, *in* ACM, ed., Proceedings of the second Nordic conference on Human-computer interaction', Vol. 31, Aarhus, pp. 101–110.

North, C. (2006), Toward Measuring Visualization Insight', *IEEE Computer Graphics and Applications* **26**(3), 6–9.

Ocker, R. & Fjermestad, J. (1998), Web-based computer-mediated communication: an experimental investigation comparing three communication modes for determining software requirements, *in* Proceedings of the Thirty-First Hawaii International Conference on System Sciences, 1998.', Kohala Coast, HI , USA.

Ogao, P. J. & Kraak, M.-J. (2002), Defining visualization operations for temporal cartographic animation design', *International Journal of Applied Earth Observation and Geoinformation* **4**(2002), 23–31.

OGC (2010), KML — OGC', http://www.opengeospatial.org/standards/kml.

Oracle (2010), MySQL :: The world's most popular open source database', http://www.mysql.com/.

Papula, L. (2001), *Mathematik für Ingenieure und Naturwissenschaftler, Band 3*, 4th edn, vieweg, Braunschweig.

Parkes, D. & Thrift, N. (1980), *Times, Spaces and Places: A Chronogeographic Perspective*, Wiley.

Pelechano, N. & Badler, N. I. (2006), Modeling Crowd and Trained Leader Behavior during Building Evacuation', *IEEE Computer Graphics and Applications* **26**(6), 80–86.

Piatti, B. (2007), A Literary Atlas of Europe - Towards a Geography of Fiction', http://www.literaturatlas.eu/index_en.html.

Piatti, B. (2008), *Die Geographie der Literatur. Schauplätze, Handlungsräume, Raumphantasien*, Wallstein, Göttingen.

Piatti, B., Bär, H. R., Reuschel, A.-K. & Hurni, L. (2008), Die Geographie der Fiktion - Das Projekt "Ein literarischer Atlas Europas"', *Kartographische Nachrichten* **58**(6), 287–294.

Pillat, R. M., Valiati, E. R. A. & Freitas, C. M. (2005), Experimental study on evaluation of multidimensional information visualization techniques, *in* CLIHC '05 Proceedings of the 2005 Latin American conference on Human-computer interaction', ACM, Cuernavaca, Mexico.

Polys, N. F. & Bowman, D. A. (2004), Design and display of enhancing information in desktop information-rich virtual environments: challenges and techniques', *Virtual Reality* **8**(1), 41–54.

Polys, N. F., North, C., Bowman, D. A., Ray, A., Moldenhauer, M. & Dandekar, C. (2004), Snap2Diverse: coordinating information visualizations and virtual environments, *in* R. F. Erbacher, P. C. Chen, J. C. Roberts, M. T. Gröhn & K. Börner, eds, Visualization and Data Analysis, Proceedings of the SPIE', Vol. 5295, SPIE, pp. 189–200.

Purves, R., Dowers, S. & Mackaness, W. (2002), Providing context in virtual reality: the example of a CAL package for mountain navigation, *in* P. Fisher & D. J. Unwin, eds, Virtual Reality in Geography', Vol. Chapter 13, Taylor & Francis, London, pp. 175–189.

Qu, H., Wang, H., Cui, W., Wu, Y. & Chan, M.-Y. (2009), Focus+Context Route Zooming and Information Overlay in 3D Urban Environments', *IEEE Transactions on Visualization and Computer Graphics* **15**(6), 1547–1554.

R (2011), The R Project for Statistical Computing', http://www.r-project.org/.

Rase, W.-D. (2003), Von 2D nach 3D - perspektivische Zeichnungen, Stereogramme, reale Modelle, *in* Kartographische Schriften, Band 7: Visualisierung und Erschliessung von Geodaten. Beiträge des Seminars GEOVIS 2003', Deutsche Gesellschaft für Kartographie, Hannover, pp. 13–24.

Ratti, C., Biderman, A., Outram, C., Britter, R., Cassi, A., Chen, X., Dunnam, J., Echeverri, P., Ingawale, M., Kardasis, A., Kang, E. R., Min, S. & Tomasinelli, M. (2010), copenhagen wheel project', http://senseable.mit.edu/copenhagenwheel/urbanData.html.

RealityMaps, D. (2010), Die Alpen in 3D - fotorealistisch und interaktiv', http://www.alpen3d.net/.

Reitinger, B., Bornik, A., Beichel, R. & Schmalstieg, D. (2006), Liver Surgery Planning Using Virtual Reality', *IEEE Computer Graphics and Applications* **26**(6), 36–47.

Rester, M., Pohl, M., Wiltner, S., Hinum, K., Miksch, S., Popow, C. & Ohmann, S. (2007), Evaluating an InfoVis Technique Using Insight Reports, *in* E. Banissi, R. A. Burkhard, G. Grinstein, U. Cvek, M. Trutschl, L. Stuart, T. G. Wyeld, G. Andrienko, J. Dykes, M. Jern, D. Groth & A. Ursyn, eds, 11th International Conference Information Visualization, IV2007', IEEE Computer Society, Zürich, Switzerland, pp. 693–700.

Risden, K., Czerwinski, M. P., Munzner, T. & Cook, D. B. (2000), An initial examination of ease of use for 2D and 3D information visualizations of web content', *Int. J. Human-Computer Studies* **53**, 695–714.

Roberts, J. C. (2008), Coordinated Multiple Views for Exploratory GeoVisualization, *in* M. Dodge, M. McDerby & M. Turner, eds, Geographic Visualization: Concepts, Tools and Applications', Wiley, Chichester, pp. 25–48.

Roberts, J. C. & Wright, M. A. E. (2006), Towards Ubiquitous Brushing for Information Visualization, *in* E. Banissi, R. A. Burkhard, A. Ursyn, J. J. Zhang, M. Bannatyne, C. Maple, A. J. Cowell, G. Y. Tan & M. Hou, eds, Information Visualization', IEEE Computer Society, pp. 151–156.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D. & van Dantzich, M. (1998), Data Mountain: Using Spatial Memory for Document Management, *in* UIST', San Francisco, CA, pp. 153–162.

Robinson, A. C., Chen, J., Lengerich, E. J., Meyer, H. G. & MacEachren, A. M. (2005), Combining Usability Techniques to Design Geovisualization Tools for Epidemiology', *Cartography and Geographic Information Science* **32**(4), 243–255.

Rock, I. (1998), *Wahrnehmung: Vom visuellen Reiz zum Sehen und Erkennen*, Spektrum Akademischer Verlag, Heidelberg.

Sandvik, B. (2010), Using Geobrowsers for Thematic Mapping', http://thematicmapping.org/.

Saraiya, P., North, C. & Duca, K. (2005), An Insight-Based Methodology for Evaluating Bioinformatics Visualizations', *IEEE Transactions on Visualization and Computer Graphics* **11**(4), 443 – 456.

Schafer, W. A. & Bowman, D. A. (2005), Integrating 2D and 3D Views for Spatial Collaboration, *in* Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work', ACM, Sanibel Island, Florida, USA, pp. 41–50.

Schnabel, M. A. (2003), 3D Maze: Creation and Translation of Abstract Virtual Architectural Environments, *in* K. T. Lee & K. Mitchell, eds, The "Second Wave" of ICT in Education: from facilitating teaching and learning to engendering education reform, the International Conference on Computers in Education (ICCE) ICCE', Hong Kong, pp. 1201–1209.

Schnabel, O. (2007), Benutzerdefinierte Diagrammsignaturen in Karten - Konzepte, Formalisierung und Implementation, PhD, ETH Zürich.

Schobesberger, D. & Patterson, T. (2008), Evaluating the Effectiveness of 2D vs. 3D Trailhead Maps - A Map User Study Conducted at Zion National Park, United States, *in* L. Hurni & K. Kriz, eds, 6th ICA Mountain Cartography Workshop', ETH Zürich, Institute of Cartography, Lenk, Switzerland, pp. 201–205.

Schratt, A. & Riedl, A. (2005), The potential of three-dimensional display-technologies for the visualization of geo-virtual environments, *in* 22nd ICA International Cartographic Conference', A Coruna.

Schütte-Krug, K. & Filli, F. (2006), Diurnal Patterns of Red Deer Cervus elaphus Activity in three Areas of the Swiss National Park, *in* F. Filli & W. Suter, eds, Huftierforschung im Schweizerischen Nationalpark / Ungulate Research in the Swiss National Park', Vol. 93, Forschungskommission des Schweizerischen Nationalparks (SCNAT Kommission), Zernez, pp. 105–116.

Seawright, J. & Gerring, J. (2008), Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options', *Political Research Quarterly* **61**(2), 294–308.

Sebrechts, M. M., Cugini, J. V., Vasilakis, J., Miller, M. S. & Laskowski, S. J. (1999), Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces, *in* Proc. 22nd Annual ACM Conference on Research and Development in Information Retrieval', pp. 3–10.

Sgrillo, R. (2010), GE-Graph: Graph for GoogleEarth', http://www.sgrillo.net/googleearth/gegraph.htm.

Shepherd, I. D. H. (2008), Travails in the Third Dimension: A Critical Evaluation of Three-dimensional Geographical Visualization, *in* M. Dodge, M. McDerby & M. Turner, eds, Geographic Visualization: Concepts, Tools and Applications', Wiley, Chichester, pp. 199–222.

Shneiderman, B. (1996), The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations', *Proc. Visual Languages 96* .

Shneiderman, B. (2003), Why Not Make Interfaces Better than 3D Reality?', *IEEE Computer Graphics and Applications* **23**(6), 12–15.

Shneiderman, B. & Plaisant, C. (2006), Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies, *in* Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization', Venice, Italy, pp. 1–7.

Siegert, F. (2010), 3D reality maps for vacation planning', *GEOconnexion International Magazine* (Dec/Jan 2010), 40–43.

Siegrist, M. (1996), The use or misuse of three-dimensional graphs to represent lower-dimensional data', *Behaviour & Information Technology* **15**(2), 96–100.

Skupin, A. & Fabrikant, S. I. (2003), Spatialization methods: a cartographic research agenda for non-geographic information visualization', *Cartography and Geographic Information Science* **30**(2), 99–119.

Slingsby, A., Dykes, J. & Wood, J. (2008), A Guide to Getting your Data into Google Earth', http://www.willisresearchnetwork.com/Lists/WRN%20News/Attachments/8/An%20introduction%20to%20Getting%20your%20Data%20in%20Google%20Earth2.pdf.

Slingsby, A., Dykes, J., Wood, J., Foote, M. & Blom, M. (2008), The Visual Exploration of Insurance Data in Google Earth, *in* D. Lambrick, ed., GIS Research UK 16th Annual Conference', Manchester, pp. 24–32.

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S. & Hedley, N. R. (2001), Cognitive and Usability Issues in Geovisualization', *Cartography and Geographic Information Science* **28**(1), 61–75.

Slocum, T. A., McMaster, R. B., Kessler, F. C. & Howard, H. H. (2005), *Thematic Cartography and Geographic Visualization*, Prentice-Hall Series in Geographic Information Science, 2nd edn, Pearson Prentice Hall, Upper Saddle River.

Smallman, H. S., St. John, M., Oonk, H. M. & Cowen, M. B. (2001), Information Availability in 2D and 3D Displays', *IEEE Computer Graphics and Applications* **Sept/Oct**, 51–57.

SNP (2010), Der Schweizerische Nationalpark im Engadin', http://www.nationalpark.ch/go/en/.

SPI (2007), SPI: the Operational Environmental Emergency Response Tool', http://eer.cmc.ec.gc.ca/index_e.php?page=s_software/spi/spi_e.html.

St. John, M., Cowen, M. B., Smallman, H. S. & Oonk, H. M. (2001), The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks ', *Human Factors: The Journal of the Human Factors and Ergonomics Society* **43**(1), 79–98.

Stanney, K. M., Mollaghasemi, M., Reeves, L., Breaux, R. & Graeber, D. A. (2003), Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design', *Int. J. Human-Computer Studies* **58**, 447–481.

Stemler, S. (2001), An overview of content analysis', *Practical Assessment, Research & Evaluation* **7**(17).

Swienty, O., Jahnke, M., Kumke, H. & Reppermund, S. (2008), Effective Visual Scanning of Geographic Information, *in* M. Sebillo, G. Vitiello & G. Schaefer, eds, Visual Information Systems: Web-Based Visual Information Search and Management', Vol. 5188 of *Lecture Notes in Computer Science*, Springer, Salerno, Italy.

swisstopo (2010), swisstopo: National Map 1:25 000', http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/maps/national/25.html.

Tao, V. & Liang, S. H. (2009), Introduction to the Spatial Sensor Web, *in* M. Madden, ed., Manual of Geographic Information Systems', ASPRS American Society for Photogrammetry and Remote Sensing, Maryland, pp. 985–1002.

Tashakkori, A. & Teddlie, C. (1998), *Mixed methodology: Combining qualitative and quantitative approaches*, Applied Social Research Methods Series (Vol. 46), Sage, Thousand Oaks, CA.

Thomas, J. J. & Cook, K. A. (2005), *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, IEEE Computer Society.

Tiede, D. & Lang, S. (2010), Analytical 3D views and virtual globes - scientific results in a familiar spatial context', *ISPRS Journal of Photogrammetry and Remote Sensing* **65**(2010), 300–307.

Tobon, C. (2005), Evaluating Geographic Visualization Tools and Methods: An Approach and Experiment Based upon User Tasks, *in* J. Dykes, A. M. MacEachren & M.-J. Kraak, eds, Exploring Geovisualization', Vol. Chapter 34, Elsevier on behalf of the International Cartographic Association, Oxford, pp. 645–666.

Tory, M., Atkins, M. S., Kirkpatrick, A. E., Nicolaou, M. & Yang, G.-Z. (2004), Eyegaze Analysis of Displays With Combined 2D and 3D Views, *in* IEEE Visualization 2005', IEEE, Minneapolis, pp. 519–526.

Tory, M., Kirkpatrick, A. E., Atkins, M. S. & Möller, T. (2006), Visualization Task Performance with 2D, 3D and Combination Displays', *IEEE Transactions on Visualization and Computer Graphics* **12**(1), 2–13.

Troyer, M. (2007), The world of H.C. Berann', http://www.berann.com/.

Tufte, E. R. (2001), *The Visual Display of Quantitative Information*, 2nd edn, Graphics Press, Cheshire.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley.

Valiati, E. R. A., Pimenta, M. S. & Freitas, C. M. (2006), A taxonomy of tasks for guiding the evaluation of multidimensional visualizations, *in* 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization', AVI, Venice, pp. 1–6.

van Wijk, J. J. (2005), The value of visualization, *in* Proceedings of IEEE Visualization 2005', IEEE Computer Society, pp. 79–86.

van Wijk, J. J. (2006), Bridging the Gaps', *IEEE Computer Graphics and Applications* **26**(6), 6–9.

VE (2007), '80% of Enterprise data has a geo component'. Says Who?', http://virtualearth.spaces.live.com/blog/cns!2BBC66E99FDCDB98!9153.entry.

Virrantaus, K., Fairbairn, D. & Kraak, M.-J. (2009), ICA Research Agenda on Cartography and GI Science', *The Cartographic Journal* **46**(2), 63–75.

W3C (2009), Scalable Vector Graphics (SVG) 1.1 Specification', http://www.w3.org/TR/2003/REC-SVG11-20030114/.

W3C (2010*a*), Extensible Markup Language (XML)', http://www.w3.org/XML/.

W3C (2010*b*), Portable Network Graphics (PNG) Specification (Second Edition)', http://www.w3.org/TR/PNG/.

W3C (2010*c*), The Extensible Stylesheet Language Family (XSL)', http://www.w3.org/Style/XSL/.

W3C (2010*d*), XML Schema', http://www.w3.org/XML/Schema.

W3D (2010), Web3D Consortium - Royalty Free, Open Standards for Real-Time 3D Communication', http://www.web3d.org/.

Wallpaper (2007), Design Awards: The Winners', http://www.wallpaper.com/interiors/design-awards-the-winners/1258.

Walsh, S. L. (2009), 35 Ways 3D is Better Than 2D - From 3D Tattoos to 3D Gaming', http://www.trendhunter.com/slideshow/3-D-pictures.

Wang, X. (2005), Integrating GIS, simulation models, and visualization in traffic impact analysis', *Computers, Environment and Urban Systems* **29**(2005), 471–496.

Wang, Y., Bowman, D., Krum, D., Coelho, E., Smith-Jackson, T., Bailey, D., Peck, S., Anand, S., Kennedy, T. & Abdrazakov, Y. (2008), Effects of Video Placement and Spatial Context Presentation on Path Reconstruction Tasks with Contextualized Videos', *IEEE Transactions on Visualization and Computer Graphics* **14**(6), 1755–1762.

Ward, M. O. (2008), Multivariate Data Glyphs: Principles and Practice, *in* C.-h. Chen, W. K. Härdle & A. Unwin, eds, Handbook of Data Visualization', Springer Handbooks of Computational Statistics, Springer, pp. 179–198.

Ware, C. (2004), *Information Visualization - Perception for Design*, The Morgan Kaufmann Series in Interactive Technologies, 2nd edn, Elsevier, San Francisco.

Ware, C. & Mitchell, P. (2005), Reevaluating stereo and motion cues for visualizing graphs in three dimensions, *in* A. I. C. P. Series, ed., 2nd symposium on Applied perception in graphics and visualization', Vol. 95, ACM Press, A Coruña, Spain, pp. 51–58.

Ware, C. & Plumlee, M. (2005), 3D Geovisualization and the Structure of Visual Space, *in* J. Dykes, A. M. MacEachren & M.-J. Kraak, eds, Exploring Geovisualization', Elsevier on behalf of the International Cartographic Association, Oxford, pp. 567–576.

Weber, A., Jenny, B., Wanner, M., Cron, J., Marty, P. & Hurni, L. (2010), Cartography Meets Gaming: Navigating Globes, Block Diagrams and 2D Maps with Gamepads and Joysticks', *The Cartographic Journal* **47**(1), 91–101.

Wehrend, S. & Lewis, C. (1990), A Problem-oriented Classification of Visualization Techniques, *in* IEEE Visualization 90', IEEE, San Francisco, pp. 139–143.

Wickens, C. D. & Baker, P. (1995), Cognitive Issues in Virtual Reality, *in* W. Barfield & T. A. Furness, eds, Virtual Environments and Advanced Interface Design', Oxford University Press, New York, pp. 514–541.

Wickens, C. D., Olmos, O., Chudy, A. & Davenport, C. (1997), Aviation Display Support for Situation Awarness, Technical report, Aviation Research Lab, Institute of Aviation, University of Illinois.

Wilkinson, A. M. (1991), *The Scientist's Handbook for Writing Papers and Dissertations*, Prentice Hall, Englewood Cliffs.

Wood, J., Dykes, J., Slingsby, A. & Clarke, K. (2007), Interactive Visual Exploration of a Large Spatio-Temporal Dataset: Reflections on a Geovisualization Mashup', *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1176–1183.

Wood, J., Kirschenbauer, S., Döllner, J., Lopes, A. & Bodum, L. (2005), Using 3D in Visualization, *in* J. Dykes, A. M. MacEachren & M.-J. Kraak, eds, Exploring Geovisualization', Vol. Chapter 14, Elsevier on behalf of the International Cartographic Association ICA, Amsterdam, pp. 295–312.

Wood, M., Pearson, D. G. & Calder, C. (2007), Comparing the effects of different 3D representations on human wayfinding, *in* G. Gartner, W. Cartwright & M. P. Peterson, eds, Location Based Services and TeleCartography', Lecture Notes in Geoinformation and Cartography, Springer, Heidelberg, pp. 345–356.

Xiang, Y., Chau, M., Atabakhsh, H. & Chen, H. (2005), Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list', *Decision Support Systems* **41**(2005), 69–83.

Yang, Y., Chen, J. X. & Beheshti, M. (2005), Nonlinear Perspective Projections and Magic Lenses: 3D View Deformation', *IEEE Computer Graphics and Applications* **25**(1), 76–84.

Yau, N. (2011), Most typical person in the world', `http://flowingdata.com/2011/03/03/most-typical-person-in-the-world/`.

Yi, J. S., Kang, Y. a., Stasko, J. & Jacko, J. (2007), Toward a Deeper Understanding of the Role of Interaction in Information Visualization', *IEEE Transactions on Visualization and Computer Graphics* **13**(6), 1224–1231.

Yin, R. K. (2003), *Case Study Research: Design and Methods*, Applied Social Research Methods, 3rd edn, Sage Publications, Thousand Oaks.

Yost, B. & North, C. (2005), Single complex glyphs versus multiple simple glyphs, *in* Conference on Human Factors in Computing Systems CHI '05', Portland, pp. 1889–1892.

Zar, J. H. (1984), *Biostatistical Analysis*, 2nd edn, Prentice Hall, New Jersey.

Zhou, M. X. & Feiner, S. K. (1998), Visual Task Characterization for Automated Visual Discourse Synthesis, *in* Computer Human Interaction CHI98', Los Angeles, pp. 392–399.

Zhu, B. & Chen, H. (2005), Using 3D interfaces to facilitate the spatial knowledge retrieval: a geo-referenced knowledge repository system', *Decision Support Systems* **40**, 167–182.

Zhu, Y. (2007), Measuring Effective Data Visualization, *in* G. Bebis, R. Boyle, B. Parvin, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Müller & T. Malzbender, eds, Advances in Visual Computing', LNCS 4842, Springer, Berlin, pp. 652–661.

Appendix A.

The Cartographic Journal article (Bleisch et al. 2008)

Bleisch, S., Dykes, J. & Nebiker, S. (2008), Evaluating the Effectiveness of Representing Numeric Information Through Abstract Graphics in 3D Desktop Virtual Environments', The Cartographic Journal 45(3), 216–226.

The Cartographic journal Vol. 45 No. 3 pp. 216–226 Geovisualisation Special Issue, August 2008

REFEREED PAPER

Evaluating the Effectiveness of Representing Numeric Information Through Abstract Graphics in 3D Desktop Virtual Environments

Susanne Bleisch, Jason Dykes and Stephan Nebiker

Institute of Geomatics Engineering, School of Architecture, Civil Engineering and Geomatics, University of Applied Sciences Northwestern Switzerland (FHNW), CH-4132 Muttenz / Switzerland
Email susanne.bleisch@fhnw.ch

The use of bars to represent numeric values in desktop virtual environments that provide information in 3D through monocular depth cues is evaluated. Using empirical experiments we test hypotheses regarding the effectiveness of participants in judging the heights of different bar combinations in four different settings (static 2D and 3D desktop virtual environment with and without frames). The results show that the participants are highly successful in identifying the taller of two bars. However, there are significant differences between the static 2D and 3D desktop virtual environment settings in terms of accuracy and task completion times when comparing bars. Characteristics such as the participant's spatial abilities or the position of the bars in the landscape do not influence the effectiveness of the judgements in our study.

INTRODUCTION

The goal of this study is to explore the degree to which bars of specific height can be utilised to represent numeric information at geographic locations in 3D desktop virtual environments. It is one of an on-going series of experiments driven by the widespread and increasing availability of data and technology through which such interactive 3D representations can be fashioned. Through these experiments we are evaluating the notion that the visual integration of 3D landscape models with the graphical representation of numeric values may, in contexts such as those specified below, provide a mechanism for the effective synthesis and graphical analysis of geographic information. The use and interpretation of 3D 'spatial bar charts' may in turn lead to faster, better informed decisions. This is dependent upon an ability to interpret and compare numeric information represented in abstract ways in 3D desktop virtual environments that use monocular depth cues such as perspective, size gradient or 'structure-from-motion' (Ware, 2004).

Landform related data is currently available from a number of diverse sources including environmental monitoring stations (e.g. Lehning, 2008), usage logs of mobile applications and specialist atlases (e.g. Piatti and Hurni, 2007). The availability of such data is likely to increase in the near future, especially with the extended utilisation of sensor networks (Botts *et al.*, 2006; Morville, 2005; Gross, 1999).

Desktop virtual reality environments or geo-browsers such as Google Earth (Google, 2008) and NASA World Wind (NASA, 2008) are hugely popular and a whole host of different data and information, such as that relating to landform, is being integrated into these interfaces to the Earth (Nebiker *et al.*, 2007; Butler, 2006). The popularity and accessibility of these desktop-based applications that are useable without any special hardware reinforces the need for research into the effectiveness of the technology (Slocum *et al.*, 2001), how such visualisation may be used for exploratory data analysis (Thomas and Cook, 2005; Gahegan *et al.*, 2001) and how we can define rules and recommendations for appropriate representations of the kind of additional information that geovisualisation relies upon in 3D environments (Polys and Bowman, 2004; Jobst and Germanchis, 2007).

Evaluating the effectiveness of combining graphical displays of numeric information with virtual landscape representations is important in this context. We do so by measuring responses to known stimuli. This is a somewhat psychophysical approach, although we do not control all parameters due to the interactive and exploratory nature of 3D desktop virtual environments. In desktop-based applications, the 3D virtual environment is projected onto the 2D screen and depth cues are used to ensure that the scene is perceived as being 3D. The most accessible and popular applications use monocular depth cues rather than

DOI: 10.1179/000870408X311404

binocular depth perception such as stereoscopic viewing (Ware, 2004) or 'True 3D' (Kirschenbauer, 2005). These have an influence on the shape and size of symbols used to represent data in the virtual environment and we analyse the way in which users are able to overcome these effects in desktop virtual environments by providing specific representations and allowing interaction with the application.

The results of this study will be used in complimentary research in which users experiment with more diverse and dense data sets and analysis tasks through which we hope to gain insight into their cognitive abilities in relating data to landform. Ultimately, we plan to involve domain experts in real applications in evaluating displays that combine abstract information with virtual realism in 3D environments. In combination, this series of studies that draw upon psychophysical, cognitive and user-centred approaches (e.g. Gilmartin, 1981, Slocum *et al.*, 2001; Fuhrmann *et al.*, 2005) will help to advance our understanding of the use of geovisualisation with 3D desktop virtual environments as an analytical device 'in-vivo' and to develop appropriate recommendations and practice.

Design of display graphics

Geo-browsers have a focus on photorealism, but may be used as a basis for exploring, analysing and comparing numeric information related to the landscape visualised (MacEachren *et al.*, 1999). In statistical graphics, it is common to use the length and area of abstract symbols to effectively show quantity (Tufte, 2001; Cleveland and McGill, 1984). Thus, typical displays for the comparison of different measurements include bar charts, spine plots and mosaic plots. Judging the height of bars positioned at different locations in two dimensional space is effectively a judgement of length or 'position [on a] non-aligned scale' (Cleveland and McGill, 1984, p. 532). Cleveland and McGill found this to be the third (length) or second (position on a non-aligned scale) most effective mode of representation for extracting quantitative information from graphs in their experiments. Thus, the application of bars differing in size may also be effective for displaying quantitative spatial information in 3D desktop virtual environments. However, in this situation a problem occurs. Changes in length and area are perceptual cues that make the viewer perceive the environment as 3D (Ware, 2004; Kraak, 1988). For example, bars representing different measurements at different locations will not only differ in height due to the value they represent but also due to the depth cue they provide according to their position (e.g. smaller in the background). In the real world, most people are aided in their judgment and perception of the environment through their ability to see in stereo. However, this ability is limited to the immediate vicinity. For objects that are more distant and sometimes also for closer objects, we rely on depth cues such as differences in size, occlusion or shading. Ware (2004, p. 289) even argues that 'stereoscopic depth can play no role at all at distances beyond 30 m'. Thus, we may be well-trained in judging objects according to depth cues rather than seeing them stereoscopically. Ware (2004, p. 262) paraphrases Hagen (1974) stating 'when we perceive pictures of objects, we enter a kind of dual perception mode. To some extent, we have a choice between accurately perceiving the size of the depicted object as though it were in a 3D space and accurately perceiving the size of the object at the picture plane.' Rock (1998) explains that in the real world the user is able to judge the size of objects (relative differences) because the position in the perceived 3D environment is known (size constancy).

We use these statements and ideas as a basis for empirical experiments investigating the judgements made by a set of users when perceiving and comparing the sizes of objects in 3D desktop virtual environments that they explore interactively. Billboards, or sprites (Akenine-Möller and Haines, 2002) of different height were inserted into a 3D environment to produce 3D bar charts in a geo-browser. These are planar objects that always face the user and thus allow navigation, but avoid further judgement difficulties such as those incurred when asking participants to consider volumes or to examine objects from different angles. The absolute width of the billboards is constant but varies visually according to the position of the bar to provide a depth cue in the virtual environment. The height of the billboard varies in relation to the quantitative value displayed and also according to the depth cue that it provides. According to Tufte's principles of graphical excellence (Tufte, 2001) data graphics should make efficient use of ink. For example, displaying one-dimensional data through areas that are two-dimensional may confuse those interpreting the graphic. In the case described here, the use of area is justified as one dimension is used as depth cue. Tufte also states that 'redundancy, upon occasion, has its uses: [...] facilitating comparison over various parts of the data [...]. (p. 98)'. To improve readability of data graphics in the 3D setting it might make sense to use a reference grid that has been shown to facilitate comparison between different panels in other contexts (Cleveland, 1994, application of Weber's Law in Baird and Noma, 1978). We compare data graphics consisting of bars of known length in static 2D and interactive 3D contexts and compare graphics that use frames of a fixed size (reference grid) with those that are frame free.

Research aims

The aim of this study is to empirically test the effectiveness and efficiency of quantitative graphics in 3D desktop virtual environments in relation to traditional static 2D quantitative graphics. Effectiveness and efficiency are measured through the accuracy of task performance and task completion time, respectively (Mackinlay, 1986). We are interested in finding out whether typical users are able to identify the differences in the heights of two bars representing different numeric values displayed in 3D desktop virtual environments and how accurately such differences can be detected. Such tasks are fundamental to the use of abstract graphics in 3D desktop virtual environments. As mentioned above, the typical user is somebody with an interest in making sense of geographically varying numeric values, such as those derived from a sensor network or from mobile applications usage logs in the context of variations in landscape form. They will have some experience of using 3D desktop virtual environments and skills and abilities in spatial reasoning and navigation. To achieve this aim, we design and implement experiments involving bar height estimation and comparison to test the following hypotheses:

- H1: Users are able to identify the taller of two bars in a 3D desktop virtual environment as well as they can in static 2D graphics.
- H2: The effectiveness of estimating differences between two bars is not significantly different in the 2D and 3D settings.
- H3: The efficiency of task completion is improved by the use of a reference grid (frame) (in the 2D and 3D settings).
- H4: The efficiency of estimating absolute values from bar lengths (with a reference grid, frame)[1] is not significantly different in the 2D and 3D settings.

The results are considered in relation to a number of factors that may influence task performance, including the positions of the bars in the landscape and the general spatial abilities of participants in the experiment (see Results section for the detailed questions).

EXPERIMENT

To explore these issues we conducted an experiment using a 2 × 2 factorial within-subject design randomising the order of the experiment assignments. The two independent variables are nature of representation (static 2D vs. 3D desktop virtual environment) and ancillary graphics (bars with and without frames). Whilst we do not isolate dimensionality from other aspects of the 3D desktop virtual environment representation (including interactivity and different information content), we refer to this representation as '3D' as our users found this a useful way of describing the distinction. Our experiment uses 20 different combinations of two values randomly selected from the range 1–99. The 20 combinations (subsequently named C1–C20) represent the range of possible bar combinations with the smaller bar varying from being between 15% of the height of the taller bar up to 97%. These 20 combinations are displayed at random locations in four different settings: the bars with frames and without frames on a surface in a 3D desktop virtual environment and as static 2D bar charts (Figure 1). The surface used in the 3D setting consists of an undulating part of the real world, made unrecognisable. We use this consistently in all 20 combinations which provide a range of comparative tasks that may be indicative of the kinds of comparisons made from a real data set. Two tasks are completed in which participants are asked to interpret numeric information from bar displays in all four settings (2D and 3D with and without frames): identifying the taller bar and comparing two bars (estimating how tall the shorter bar is compared with the taller). Additionally, the task of estimating absolute bar heights from bar displays with frames in the 2D and 3D setting was completed. This allows us to evaluate the participant's accuracy in estimating the height of a single bar in contrast to the task of comparing the two bars.

[1]Note that absolute height judgements of bars are only possible with a reference grid. The bars without frames could only be judged absolutely by providing further scale references or functionality.

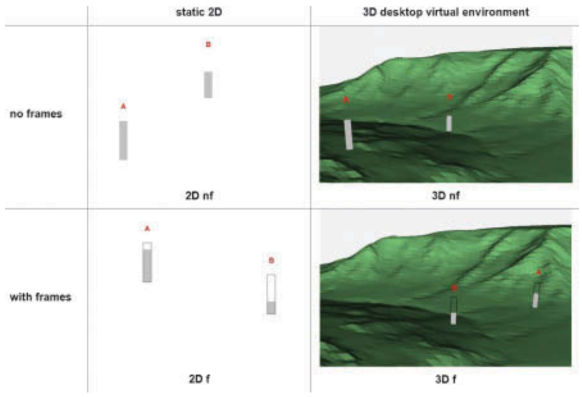


Figure 1. Four different settings of the 2 × 2 factorial design (static 2D representations without '2D nf' and with frames '2D f' and interactive 3D desktop virtual environments containing bars without '3D nf' and with frames '3D f'), see Figure 13 and Figure 14 for all bar combinations C1–C20

Participants and implementation

The experiment was conducted with a group of 26 final semester bachelor geomatics students who had some experience of using 3D displays. Each participant completed each task with a random selection of 10 of the 20 bar combinations in each setting ('2D f', '2D nf', '3D f' and '3D nf'). A subgroup of 18 participants conducted the additional experiment of estimating absolute bar heights in the '2D f' and '3D f' settings. The information was displayed using JPEG images (2D) and X3D environments (3D). The 3D environments were viewed and navigated using the Flux Player software. The participants were encouraged to navigate in the virtual 3D environment if they thought it helped accomplish the tasks but were reminded that this was an efficiency test and that completion times were being recorded. The experiments in 2D and 3D were administered and the participant's answers and task times were recorded using the quiz facility of the WebCT e-Learning platform with which participants were familiar. They performed the experiment on generally available desktop computers at our institution in controlled and consistent conditions. After performing the different tasks, the participants were given the option of commenting in writing on any aspect of the experiment and their performance. This helps us to triangulate between the subjective opinions and ideas of the participants and the more formal numeric data analysis (Elmes *et al.*, 2006).

Collected data

The data collected for each task included a statement regarding which of the two bars participants judged to be taller (A, B or equal), the percentage value when estimating the height of the shorter bar in comparison to the taller bar and the time needed to fulfil those two tasks (judging the displays and recording the answers). For the 3D settings the time recorded also includes the duration of starting and closing the 3D scene, which takes an average of 3 s but depends on the load on the Internet connection. The

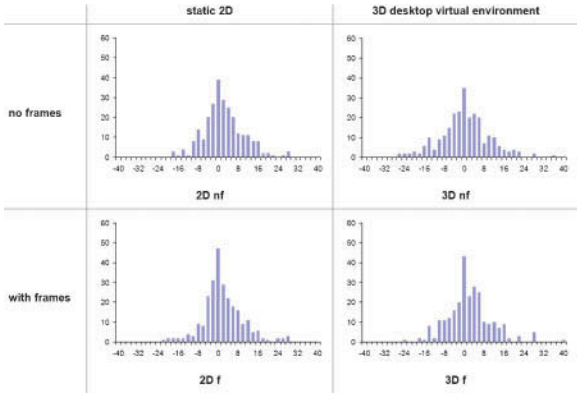


Figure 2. Histograms of the differences of the bar comparison in all four settings (class width = 2, n=260 per setting). Differences are normally distributed

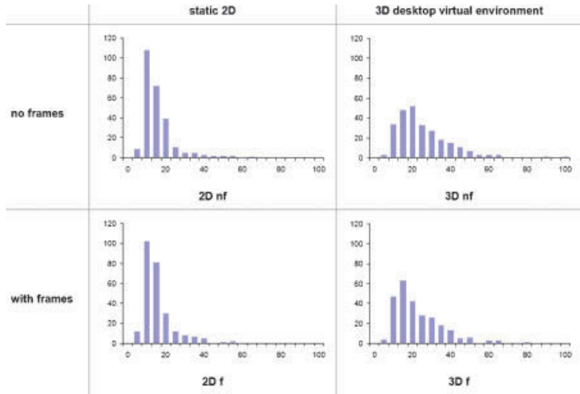


Figure 3. Histograms of task completion times in all four settings (class width = 5 s, n=260 per setting), 3 s of 3D scene start-up time were subtracted in the two 3D settings '3D f' and '3D nf'. Task completion times are lognormally distributed

accuracy of the comparison of the two bars in each setting is evaluated using the difference between the participant's estimated value and the actual percentage value (e.g. smaller bar is 67% of the taller bar, if the participant's estimation is 70%, then the difference of +3 is used for the evaluation of the results). The expected values for these differences are 0 for precise estimations by the participants. There is a slight bias in the data as people tend to estimate the differences in 5 or 10 s (e.g. estimation of 65 or 70% and not 67%). But the data compensates for this bias as under and over judgments can be assumed to be made with equal frequency.

Three of the 1400 data sets collected were ignored because of null responses. The data were checked for errors and inconsistencies, such as participants using 0 instead of 100 to record equality or entries using decimals instead of percentages for comparison. Such cases were corrected before analysis.

For the analysis of the estimation of the absolute bar height in the 2D and 3D settings with frames the differences between the estimated and the actual bar heights are used. The expected value for this difference is 0. Figure 4 shows the histograms of the differences in the 2D and 3D setting. From the 360 data sets collected, 24 were ignored as in these cases participants compared the two different bars rather than estimating the absolute heights of the two bars. The data set sizes are n=179 for '2D f' and n=157 for '3D f'.

Data analysis

The quantitative data were statistically analysed to test our hypotheses (see Research aims subsection above). The data sets approximate to a normal distribution (differences – shown in Figure 2 and Figure 4) or lognormal distribution (times – shown in Figure 3) and so it is appropriate to describe them by their means, standard deviations and minimum and maximum values. Different aspects of the data are compared and related to influencing factors by calculating correlations between the data sets.

Student t-tests were employed for hypothesis testing, using a significance level of 95% (α = 0.05), to test for differences of the mean from 0. Differences between the data sets of the different settings are evaluated using ANOVA and Tukey-Tests (Zar, 1984).

RESULTS

H1: Identification of the taller bar

The analysis shows that the participants were able to identify the taller bar in all four settings (2D and 3D with and without frames) in almost 100% of the cases (Figure 5). In each setting, a number of bar combinations, mostly with proportions higher than 80%, were judged as being equal (Figure 6, and shown by the grey portions in Figure 5).

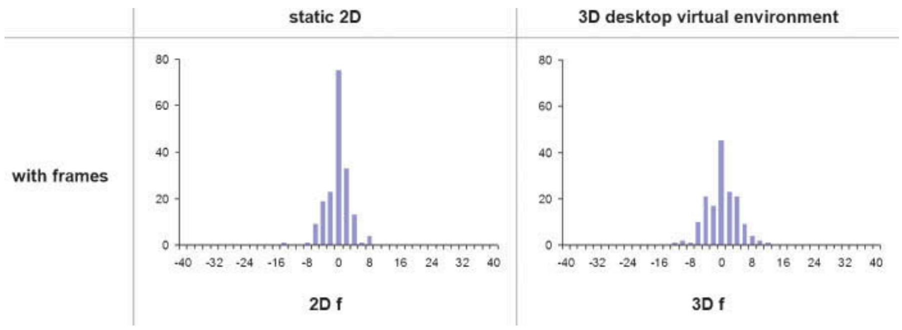


Figure 4. Histograms of the differences between estimated and absolute values of bar heights in the '2D f' and '3D f' settings. Differences are normally distributed

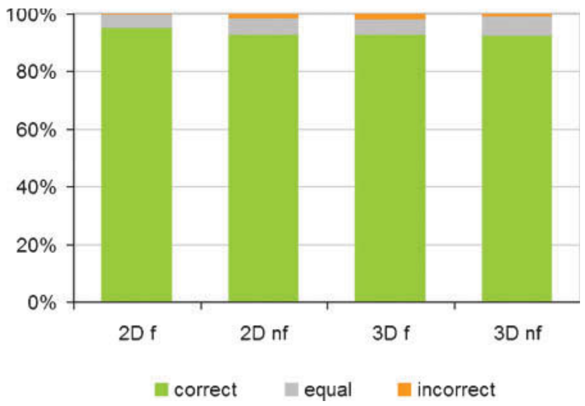


Figure 5. Correct and incorrect judgments of the taller bar. Grey values indicate bar combinations judged as being of equal size. (n=260 per setting)

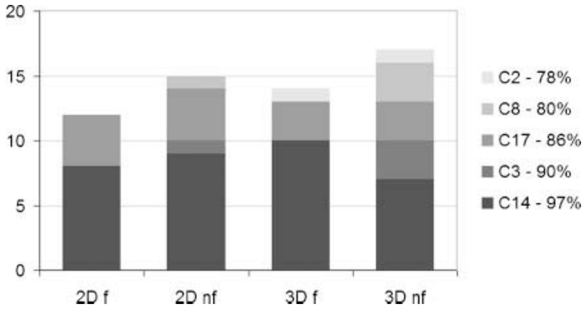


Figure 6. Frequency of bar combinations judged as being of equal size (proportion of the two bars as a percentage)

The bar combinations for which the incorrect bar was selected as being the taller one are shown in Table 1 for each setting. These combinations are equivalent to the orange portions of Figure 5. The percentage values show the height of the smaller bar in proportion to the larger. Images of all bar combinations (C1–C20) in the 2D and 3D settings can be found in Figure 13 and Figure 14.

H2 and H3: Estimating the difference between two bars

The task of comparing the absolute values of the two bars (Figure 7 and Figure 8) is completed with the same level of success in the 3D settings as in the 2D settings. The frames do not improve the accuracy of the comparison but the participants are a little faster and seven qualitative statements mention that the tasks feel easier to complete with frames. Two different statements note that for the settings with frames (2D and 3D) the task of comparing the two bars is more complex. Rather than comparing the two bars directly participants initially estimated the height of each

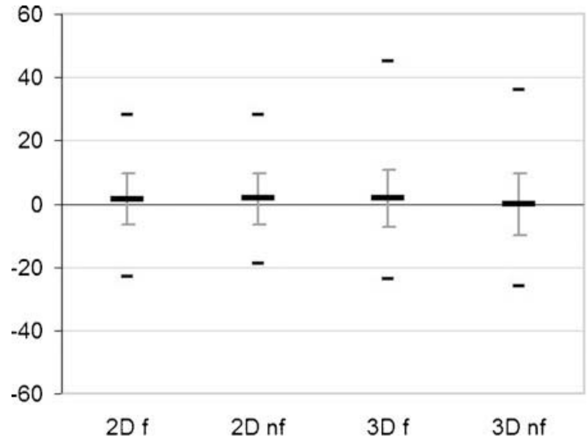


Figure 7. Mean, standard deviation, and minimum/maximum values of the differences between estimated and actual values in all four settings (n=260 per setting)

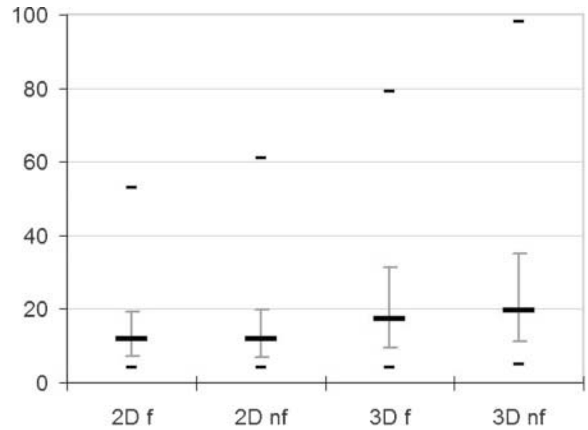


Figure 8. Mean, standard deviation (calculated from lognormal distribution), and minimum/ maximum values of task completion time in seconds in all four settings (n=260 per setting), 3 s of 3D scene start-up time were subtracted in the two 3D settings '3D f' and '3D nf'

bar and then performed the comparison from those values. Despite this reported complexity these participants are more content with task completion in the settings with frames.

Applying a Student t-test shows that the means of the differences between estimated and actual values for the settings '2D f', '2D nf' and '3D f' (Figure 7) are significantly different from 0 (α = 0.05). The mean of the differences in setting '3D nf' is not significantly different from 0. Using ANOVA and Tukey-Tests to test the hypothesis that the means of the differences per setting are equal (H0: $m_{2Df} = m_{2Dnf} = m_{3Df} = m_{3Dnf}$) results in the rejection of H_0. The alternative hypothesis

Table 1. Selection of incorrect bar as being the higher bar for each setting (proportion of the two bars as a percentage)

2D f	2D nf	3D f	3D nf
1 × C19 – 22%	1 × C6 – 50% 1 × C13 – 64% 1 × C10 – 66% 1 × C1 – 68%	1 × C9 – 15% 1 × C6 – 50% 1 × C5 – 63% 1 × C3 – 90% 1 × C14 – 97%	1 × C3 – 90% 2 × C14 – 97%

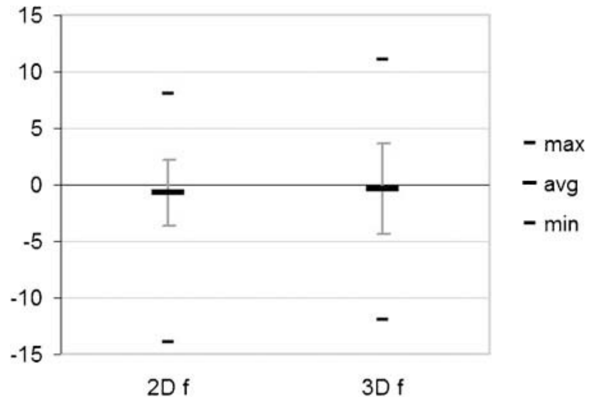


Figure 9. Mean, standard deviation, and minimum/maximum values of the differences between recorded and absolute values when estimating absolute values

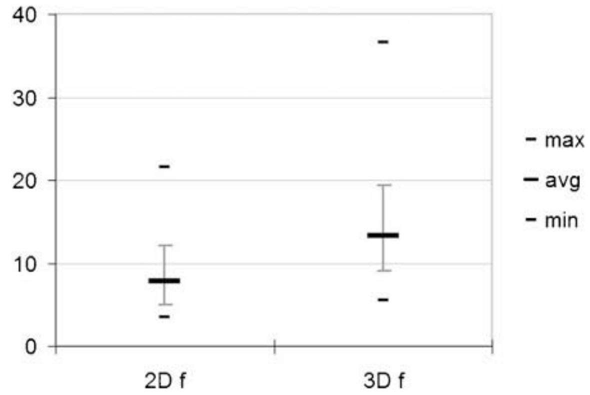


Figure 10. Mean, standard deviation (calculated from lognormal distribution), and minimum/ maximum values of tasks completion time for estimating and recording absolute values

H1: $m_{2Df} = m_{2Dnf} = m_{3Df} \neq m_{3Dnf}$ is accepted at $\alpha = 0.05$. This analysis allows us to conclude that participants in our experiment tended to slightly, but significantly, overestimate the size of the smaller bar in respect to the taller bar in the two 2D settings and in the '3D f' setting, but not in the case of '3D nf'.

Using ANOVA and Tukey-Tests to test the hypothesis that the means of the times per setting are equal (H0: $m_{2Df} = m_{2Dnf} = m_{3Df} = m_{3Dnf}$) results in the rejection of H_0. Alternatively, the hypothesis H1: $m_{2Df} = m_{2Dnf} \neq m_{3Df} \neq m_{3Dnf}$ is accepted (note that 3 s of 3D scene start-up time were subtracted in the two 3D settings '3D f' and '3D nf'). Estimating the differences in size between the two bars takes significantly more time in the 3D settings and within the 3D settings significantly more time for the '3D nf' setting. The qualitative statements show that some participants felt they spent a considerable amount of time navigating the 3D scenes. Whilst navigation, exploration and spending time with the scene content may reduce efficiency as measured here it may have benefits that we are not measuring in these experiments such as increased confidence in results or understanding of topography and the spatial arrangement of the measurements. Such aspects will be the focus of future experiments.

H4: Estimating absolute bar heights

The task of estimating absolute bar heights from displays with frames is completed with the same level of success in the 3D setting as in the 2D setting (Figure 9 and Figure 10). Applying a Student t-test shows that the mean of the differences (Figure 9) is significantly different from 0 for the setting '2D f' ($\alpha = 0.05$). Here, the participants tend to slightly underestimate the bar heights when responses are compared with the actual values. Nevertheless, the high accuracy of bar height estimation especially in the '3D f' setting helps explain why the task of comparing two bars in the settings with frames is still highly accurate even though participants stated that they tend not to compare the bars directly but rather estimate the heights independently first before undertaking the comparison.

Using a Tukey-Test for the means of the task completion times, the alternative hypothesis $m_{2Df} \neq m_{3Df}$ is accepted with the 3D setting once again taking longer (note that 3 seconds of 3D scene start-up time were subtracted in the setting '3D f').

Influence of general spatial abilities of the participants

The results gained from participants in the two 3D settings are compared with general spatial abilities as measured by the Santa Barbara Sense of Direction test (Hegarty, Richardson *et al.*, 2002) to provide some insights into one of a number of possible influences. The SBSOD test was conducted with a subgroup of 10 participants, all of whom achieved a reasonably high sense of direction (SOD) value, confirming that our efforts to focus on users with spatial and navigation skills had been successful. This small number, however, means that any results can only be used to suggest trends. The correlation coefficients r (Table 2) and scatter plots (Figure 11) do not strongly suggest that correlations exist between the SOD of the participants and their task performance in the two 3D settings when measured using absolute means of differences between estimated and actual values of bar size differences. The task performance in one 3D setting is, however, strongly correlated ($r = 0.727$) with the task performance in the other 3D setting (comparing the means of differences per participant) showing some consistency in the performance and response of individual users.

Influence of various aspects on the experiment outcomes

Various different aspects of the experiments conducted here may have an influence on the results reported above. We calculated correlations to answer the following questions that relate to the geography of our examples and the time that users spent on the tasks:

Table 2. Correlation coefficients (r) describing the relationships between sense of direction (SOD) of participants and the participant's task performance (absolute means of differences) in the two 3D settings and between the two 3D settings (with and without frames, means of differences)

	3D nf	3D f
SOD	−0.035	0.376
3D f	0.727	–

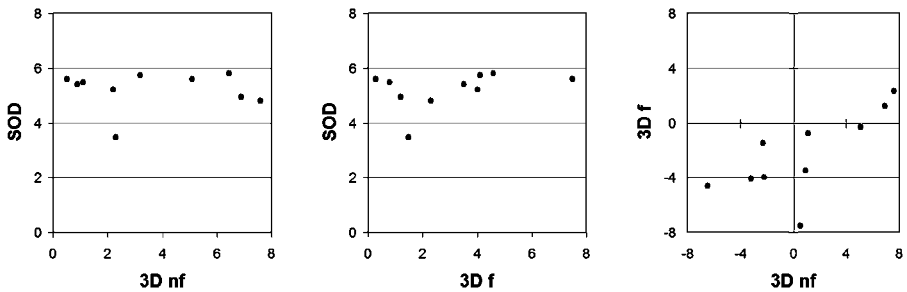


Figure 11. Scatter plots of the relationships between sense of direction (SOD) of participants and the participant's task performance (absolute means of differences) in the two 3D settings and between the two 3D settings (with and without frames, means of differences)

1) Do participants perform better when they spend more time on the different tasks?
2) 2D settings: Are bar combinations easier to evaluate when more closely aligned along a common base line?
3) 3D settings: Are bar combinations more accurately compared if they are nearer to each other in the landscape?

1) Coefficients of correlation between the means of completion times and the absolute means of the bar comparison differences were calculated for each participant and setting. Table 3 shows that task performance does not seem to be correlated with the time taken to perform the tasks. Participants may need more time to fulfil the tasks but are not performing more effectively in terms of our metrics.

2) We calculated coefficients of correlation between absolute means and the standard deviations of the bar comparison differences per bar combination in the two 2D settings with values describing the absolute vertical non-alignment of the bar combinations, the horizontal distance between bars and the 2D diagonal distance between the two bars. Table 4 shows that the performance per bar combination (measured as the absolute means of the differences per bar combination) in the two 2D settings is not correlated with the absolute vertical bar non-alignment. The standard deviations of the differences show a weak positive correlation with the absolute vertical bar non-alignment. This suggests that the bar comparison results may vary more the further the two bars are vertically separated coinciding to some degree with the findings of Cleveland and McGill (1984) regarding non-alignment in the 2D setting.

3) We correlated the absolute means and the standard deviations of the differences of bar comparison in the 3D settings with the 3D distance (calculated from horizontal and vertical differences between the two bars) between the bars per bar combination. Table 5 shows that the performance per bar combination (measured as the absolute means of the differences per bar combination) in the two 3D settings is not correlated with the 3D distance between the two bars. The standard deviations of the differences are to some degree correlated with the 3D distance, though this is a weak correlation. This suggests that here, as in the 2D settings, bigger vertical or horizontal differences between the two bars may result in higher variation in the results but there is no evidence that it results in less accuracy. Similar experiments may be conducted in areas were the topography varies more dramatically. The effects of horizontal, vertical or 3D distances between bar locations may be more influential in such

Table 3. Correlation coefficients for the correlation between the participant's mean task performance time and their absolute mean of bar comparison differences per setting

Absolute mean of differences	2D f	2D nf	3D f	3D nf
Mean times	−0.198	−0.261	−0.181	−0.153

Table 4. Correlation coefficients for the correlation between the means and standard deviations of differences of bar comparisons and the absolute vertical non-alignment per bar combination, the horizontal distance and the diagonal distance

Correlation between…	Absolute means of differences	Standard deviations of differences
Absolute vertical bar non-alignments	0.178	0.365
Horizontal distances between bars	0.043	−0.119
Diagonal distances between bars	0.060	−0.054

Table 5. Correlation coefficients for the correlation between the mean of differences of bar comparison per bar combination and the 3D distance between the two bars per bar combination

Correlation between…	Absolute means of differences	Standard deviations of differences
3D distances between bars	0.084	0.249

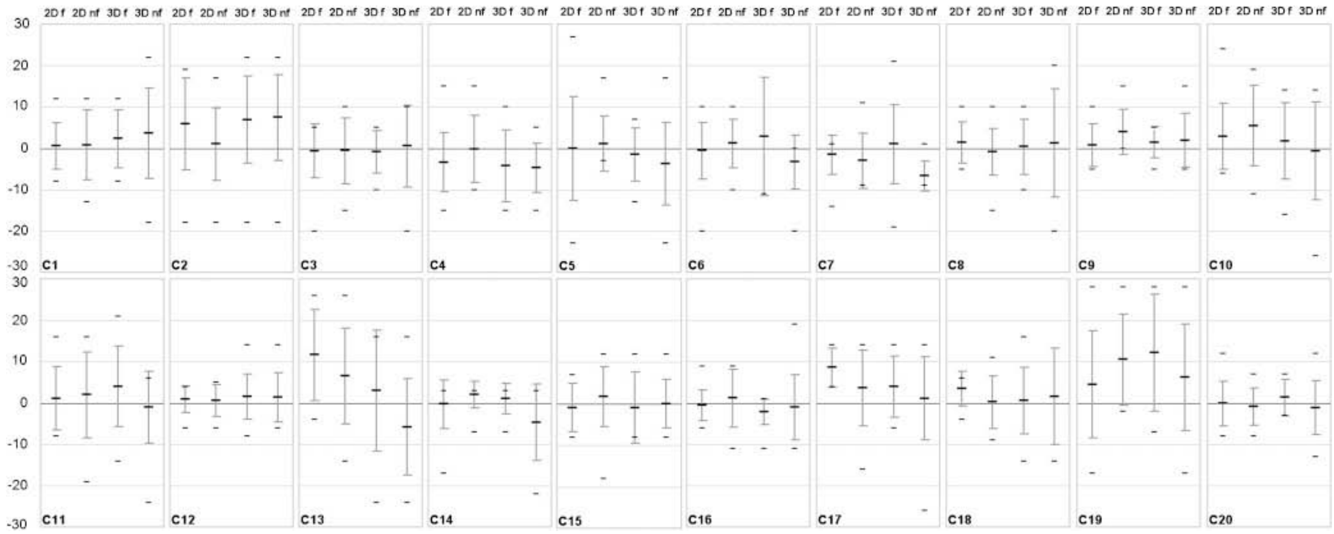


Figure 12. Mean, standard deviation, and minimum/maximum values of bar comparison differences per bar combination (C1–C20) for all four settings ('2D f', '2D nf', '3D f' and '3D nf')

landscapes as may be the case in sections of a landscape in which occlusion is an issue.

Comparing the different bar combinations

The same 20 combinations of bars (C1–C20) are used for the different tasks in all four of our settings (2D and 3D with and without frames). Thus, the performance per bar combination can be evaluated and compared between the different settings. Figure 12 shows the participants' performance per bar combination in all four settings ('2D f', '2D nf', '3D f' and '3D nf'). The results shown here can only suggest trends as the sizes of the datasets per bar combination and setting are small ($8 \leqslant n \geqslant 18$). Certain bar combinations, seem more difficult to compare in all four settings than others. For example, performance is strong with combinations C9, C16 or C20 and less strong with C2, C10 and C13, which seem more difficult combinations (images of the bar combinations in the 2D and 3D settings are shown in Figure 13 and Figure 14). This suggests that small bars with little difference are more difficult to compare than taller bars or bars with a big difference in height. In practice, we are unlikely to be able to freely define the size of the bars that represent real measurements recorded in the field. However, appropriate scaling of the data may improve the readability and comparison of data displays where bars are small and these results support the case for interactive tools that aid comparison between bar values in certain cases.

CONCLUSIONS AND IMPLICATIONS

The results reported here show that when pairs of bars of varying height and constant width are used to represent numeric values in 3D desktop virtual environments, the numbers represented can be estimated and compared effectively. Skilled users are very successful in separating the perception of monocular depth cues such as perceived variations in the width and height of the bars in the landscape from the actual values the bars represent by their heights. It seems that we are indeed well-trained in judging objects according to depth cues rather than seeing them stereoscopically as indicated in the Introduction. This knowledge serves as a basis for studies evaluating more dense and multivariate data sets with more complex tasks in virtual 3D environments.

Research Hypotheses

Revisiting the research hypotheses we find the following:

H1: Users are able to identify the taller of two bars in a 3D desktop virtual environment as well as they can in static 2D graphics.

Informally, the evidence suggests that the participants are able to identify the taller of two bars in all four settings. Thus hypothesis 1 is accepted. As we have shown above, the taller bar could be identified in

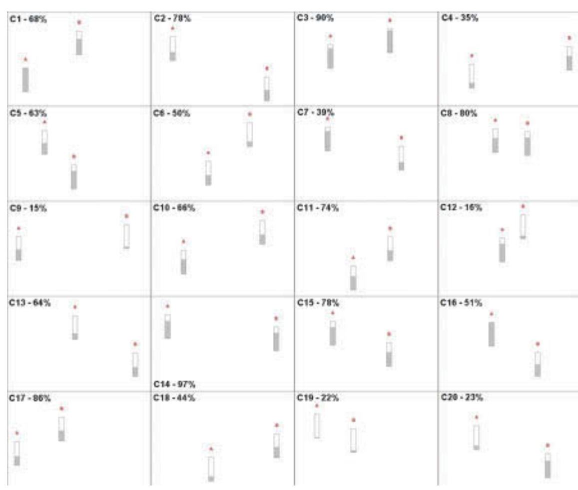


Figure 13. All 20 bar combinations in the '2D f' setting. The proportion of the smaller bar to the taller bar is given as a percentage ('2D nf' setting: same bar positions but without frames)

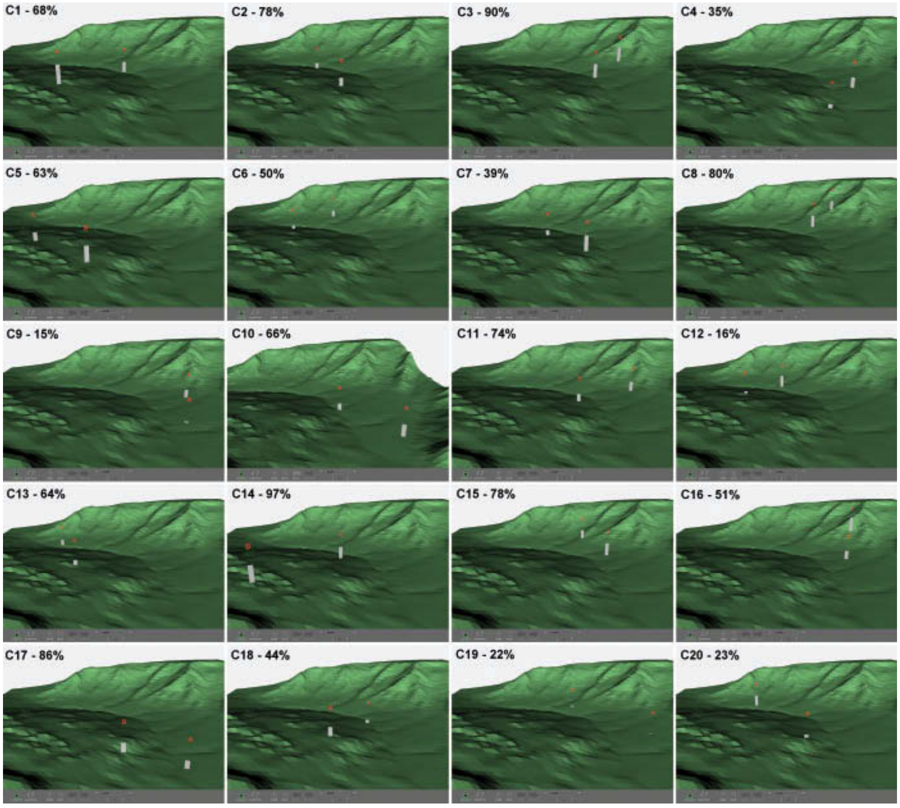


Figure 14. All 20 bar combinations in the '3D nf' setting (screenshots of explorable 3D visualisations from a fixed viewpoint). The proportion of the smaller bar to the taller is given as a percentage ('3D f' setting: same bar positions but with frames)

almost 100% of the cases in all four settings ('2D f', '2D nf', '3D f' and '3D nf'). Most errors occurred where the smaller bar was sized 86% or more of the larger bar, independent of absolute bar size. Even though the identification of the tallest bar in a display is easily computed and indicated, for example by a different colour, it is assumed that the successful completion of this basic task is helpful for mastering more complex tasks such as identifying clusters of bars that are taller than the surrounding bars or relating bar heights to landform.

H2: *The effectiveness of estimating differences between two bars is not significantly different in the 2D and 3D settings.*
The results discussed above show that we reject hypothesis 2 as there are significant differences between the 2D and 3D settings.

H3: *The efficiency of task completion is improved by the use of a reference grid (frame) (in the 2D and 3D settings).*
Hypothesis 3 is rejected for the 2D settings as there were no significant differences found between the settings with and without frames in 2D. However, in the 3D setting with frames the tasks take significantly less time than in the 3D setting without frames.
In general, participants tend to slightly overestimate the size of the smaller bar in respect to the taller bar. However, the use of frames does not have significant effects on the accuracy of the judgments of bar differences in either positive or negative ways. The task of comparing bar heights took significantly more time in the 3D settings compared with the 2D settings and was thus less efficient. This effect may be attributed to the fact that the participants navigated in the 3D scenes before making a judgement of bar heights. Ware (2004) mentions that no interaction or navigation can ever be as fast as simple eye movement such as needed for the 2D settings and our findings support this contention. Further experiments will need to evaluate whether the additional efforts in terms of time are worthwhile compared with possible gains that may be achieved when judging numeric values in the context of 3D landscapes.

H4: *The efficiency of estimating absolute values from bar lengths (with a reference grid, frame) is not significantly different in the 2D and 3D settings.*
The results discussed above require that hypothesis 4 is rejected. There were significant differences found between the 2D and 3D settings with frames. However, the results are very accurate and the differences in time between the two settings may be mainly attributed to the navigation in the 3D space.

Influencing Factors

We also consider the influences of recorded characteristics that may affect performance. These include the effects of participants' general spatial ability on task performance and the task completion times, the position of the bars in the landscape, the vertical non-alignment of the bars in the 2D settings and the different combinations of bars.

Spatial abilities of the participants: The Santa Barbara Sense of Direction test allowed us to confirm that our participants were characteristic of the 'typical users' in whom we have interest – assuming that this is a suitable measure of spatial abilities. However, our analysis does not allow us to establish this factor as an influence on task performance. Some consistency in the performance and response of individual users is shown as the task performance in one 3D setting is highly correlated with the task performance in the other 3D setting. Further research may use more discriminating tests within our target user-group.

Task completion times: There is no significant evidence of a correlation between task performance and task completion time in either of the settings. Participants taking more time to complete a task may have more difficulties with the task or be participating in additional activity.

Vertical non-alignment of bars in the 2D settings: There is no evidence of a correlation between accuracy of task performance per bar combination in the 2D settings and the absolute vertical non-alignment of the bars. However, the standard deviation of the differences exhibits a weak correlation with non-alignment suggesting that the results vary more when bars are not aligned along a common baseline. This is to some degree consistent with the findings of Cleveland and McGill (1984) who note that judging positions along a common scale is easier than judging positions along non-aligned scales when considering 2D graphics.

Position of bars in the 3D landscape: The position of the bars in the landscape (3D distance between bars) does not seem to influence the accuracy of task performance. But as is the case for non-alignment in the 2D setting, the 3D distance between the bars shows a weak positive correlation with the standard deviation of the differences. Whilst this is not a significant finding it suggests that greater distances between two bars may result in variability in the judgment of results. There may be an effect relating to the geography of data analysis in virtual worlds whereby comparisons between closer symbols are more efficient than comparisons between distant symbols. Further research may need to consider the effects of this issue when more than two bars are under consideration concurrently.

Bar combinations: Certain bar combinations, especially combinations of small bars with minor differences in height seem more difficult to compare (in all four settings) than others. However, this is only a trend as the data set sizes per bar combination are small ($8 \leqslant n \geqslant 18$). In practice, such knowledge may help to appropriately scale the data to improve the readability and comparison of data displays where bars are small or at least support the need to provide interactive tools that aid comparison between bar values.

Concluding Comments

These results provide some insights into the effectiveness of using abstract symbols to represent numeric data values in 3D desktop virtual environments or geo-browsers and associated efficiencies and lead to a number of inferences. It is suggested that the key difference between the 3D and 2D settings is in task completion time (efficiency) rather than effectiveness. However, it is difficult to isolate dimensionality, interactivity or information content when making this conclusion. It may be that tasks that justify 3D desktop virtual environments involve the assimilation and visual synthesis of additional (spatial) information that cannot be represented effectively in 2D. Such tasks may be more complex and justify longer completion times. Further studies will focus on measuring possible additional benefits when judging and relating abstract symbols to the landscape by navigation and interaction which may make the additional efforts in terms of time worthwhile. Our results indicate that fundamental estimation and comparison tasks – the foundation of data graphics – are not affected by a combination of factors such as monocular depth cues or interactivity which contribute to 3D desktop virtual environments.

Using bars with frames appears to help users in 3D desktop virtual environments. They are more confident in their judgments and also faster in doing so in the virtual environment. Even though some participants state that it is more complex, the completion times show that frames seem to reduce the need to navigate or explore the virtual environment. This is somewhat backed-up by a single statement noting that more navigation, especially zooming in, was needed for the virtual environment setting without frames and this had an effect on task completion times. Frames are one of a number of additional graphical and or numeric devices that might improve the efficiency of task completion. Developers using 3D desktop virtual environments to combine landscape information with statistical graphics may want to consider using some kind of reference grid, such as the frames in this study, for their displays. We will test in subsequent evaluations if this recommendation holds true for more dense displays and more complex analysis tasks.

Relating our findings back to the rationale for the study, we conclude that we are not yet in the position to know whether the use and interpretation of 3D 'spatial bar charts' in 3D desktop virtual environments leads to faster and/or better informed decisions. However, the results of this study indicate that our users can perform basic estimation and comparison tasks when interacting with 3D scenes that use monocular cues to represent depth. We are in the process of extending our knowledge through complimentary research in which users experiment with more diverse and dense data sets and analysis tasks. Ultimately, we plan to evaluate displays that combine abstract information with virtual realism in 3D environments with domain experts in real world applications. The results generated by the experiments reported here employ a somewhat psychophysical approach to form the foundations for subsequent experiments with more cognitive and user-centred perspectives. In combination, this series of studies will help to advance our understanding of the use of geovisualisation with 3D desktop virtual environments as an analytical device 'in-vivo', to learn about the implications on decision making and to develop appropriate recommendations and practice.

BIOGRAPHICAL NOTES

Susanne Bleisch is a PhD candidate at the giCentre, City University London and a scientific collaborator for the e-learning projects CartouCHe (3D cartography), GITTA (Geographic Information Technology and Training Alliance) and eLML (eLesson Markup Language) at the Institute of Geomatics Engineering of the University of Applied Sciences Northwestern Switzerland – FHNW. She has studied Geomatics at FHNW in Muttenz, Switzerland and obtained her Masters degree in Geographic Information from City University in London. Her research interests are in 3D geovisualisation, cartography and geoinformatics and e-learning.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Dr. Beat Fischer from the University of Applied Sciences Northwestern Switzerland FHNW in Muttenz for helpful discussion and comments related to the statistical analysis of the data. We are grateful to the test participants for their willingness to participate and their efforts in judging the different displays. The authors also thank the three anonymous reviewers of this paper for their comments and suggestions which have been used to improve the work.

REFERENCES

Akenine-Möller, T. and E. Haines (2002). **Real-Time Rendering**. Natick, A K Peters.

Baird, J. C. and E. Noma (1978). **Fundamentals of Scaling and Psychophysics**. New York, Wiley.

Boots, M., Percivall, G., Reed, C. and Davidson, J. (2006). OGC Sensor Web Enablement: Overview and High Level Architecture. OGC Whitepaper.

Butler, D. (2006). 'The web-wide world.' **Nature** 439(16): 776–778.

Cleveland, W. S. and R. McGill (1984). 'Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods.' **Journal of the American Statistical Association** 79 (387): 531–554.

Cleveland, W. S. (1994). **The Elements of Graphing Data**. New Jersey, Hobart Press.

Elmes, D. G., Kantowitz, B. H., and Roediger, H. L. (2006). **Research Methods in Psychology**. Belmont, Thomson Wadsworth.

Fuhrmann, S., Ahonen-Rainio, P., Edsall, R. M., Fabrikant, S. I., Koua, E. L., Tobon, C., Ware, C. and Wilson, S. (2005). Making Useful and Useable Geovisualisation: Design and Evaluation Issues. **Exploring Geovisualisation**. J. Dykes, A. M. MacEachren and M.-J. Kraak. Amsterdam, Elsevier, pp. 553–566.

Gahegan, M., Wachowicz, M., Harrower, M. and Rhyne, T.-M., (2001). 'The Integration of Geographic Visualisation with Knowledge Discovery in Databases and Geocomputation.' **Cartography and Geographic Information Science** 28(1): 29–44.

Gilmartin, P. P. (1981). 'The Interface of Cognitive and Psychophysical Research in Cartography.' **Cartographica** 18(3): 9–20.

Google. (2008). Google Earth – Explore, Search and Discover. Retrieved 12 Jan, 2008, from http://earth.google.com.

Gross, N. (1999). The earth will don an electronic skin. **BusinessWeek**.

Hegarty, M., A. E. Richardson *et al.* (2002). 'Development of a self-report measure of environmental spatial ability.' **Intelligence** 30: 425–447.

Jobst, M. and Germanchis, T., (2007). 'The Employment of 3D in Cartography – An Overview.' **Multimedia Cartography**. W. Cartwright, M. P. Peterson and G. Gartner. Berlin, Springer, pp. 217–228.

Kirschenbauer, S. (2005). 'Applying "True 3D" Techniques to Geovisualisation: An Empirical Study.' **Exploring Geovisualisation**. J. Dykes, A. M. MacEachren and M.-J. Kraak. Amsterdam, Elsevier, pp. 363–387.

Kraak, M.-J. (1988). Computer-assisted Cartographical Three-dimensional Imaging Techniques. Delft, Delft University Press.

Lehning, M. (2008). Wannengrat field test site. Retrieved 14 March 2008, from http://www.wsl.ch/forschung/forschungsunits/schnee/Schneedecke_und_Mikrometeorologie/Wind-flow-transportation-and-surface-interactions/Wannengrat/index_EN?redir=1&

MacEachren, A. M., Edsall, R., Haug, D., Baxter, R., Otto, G., Masters, R., Fuhrmann, S. and Quian, L. (1999). Virtual environments for geographic visualisation: potential and challenges. Proceedings of the ACM Workshop on New Paradigms in Information Visualisation and Manipuation. Kansas City, pp. 35–40.

Mackinly, J. (1986). 'Automating the design of graphical presentations of relational information.' **ACM Transactions on Graphics** 5(2): 110–141.

Morville, P. (2005). **Ambient Findability**. Sebastopol, CA, O'Reilly.

Nebiker, S., Christen, M., Eugster, H., Flückiger, K. and Stierli, C. (2007). Integrating mobile geo sensors into collaborative virtual globes – design and implementation issues. Proceedings of the 5th International Symposium on Mobile Mapping Technology MMT'07, Padua.

NASA. (2008). NASA World Wind. Retrieved 12 Jan, 2008, from http://worldwind.arc.nasa.gov.

Piatti, B. and L. Hurni (2007). Towards a European Atlas of Literature: Developing Theories, Methods, and Tools in the Field of 'Literary Geography'. XXIII International Cartographic Conference ICC, Moscow.

Polys, N. F. and Bowman, D. A., (2004). 'Design and display of enhancing information in desktop information-rich virtual environments: challenges and techniques.' **Virtual Reality** 8(1): 41–54.

Rock, I. (1998). **Wahrnehmung: Vom visuellen Reiz zum Sehen und Erkennen**. Heidelberg, Spektrum Akademischer Verlag.

Slocum, T. A.; Blok, C.; Jiang, B.; Koussoulakou, A.; Montello, D. R.; Fuhrmann, S. and Hedley, N. R.. (2001). 'Cognitive and Usability Issues in Geovisualisation.' **Cartography and Geographic Information Science** 28(1): 61–75.

Thomas, J. J. and Cook, K. A., (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics, IEEE Computer Society.

Tufte, E. R. (2001). **The Visual Display of Quantitative Information**. Cheshire, Graphics Press.

Ware, C. (2004). Information Visualisation – Perception for Design. San Francisco, Elsevier.

Zar, J. H. (1984). **Biostatistical Analysis**. New Jersey, Prentice Hall.

Appendix B.

Data sets and visualisations

B.1. Stage IIa: data sets and visualisations

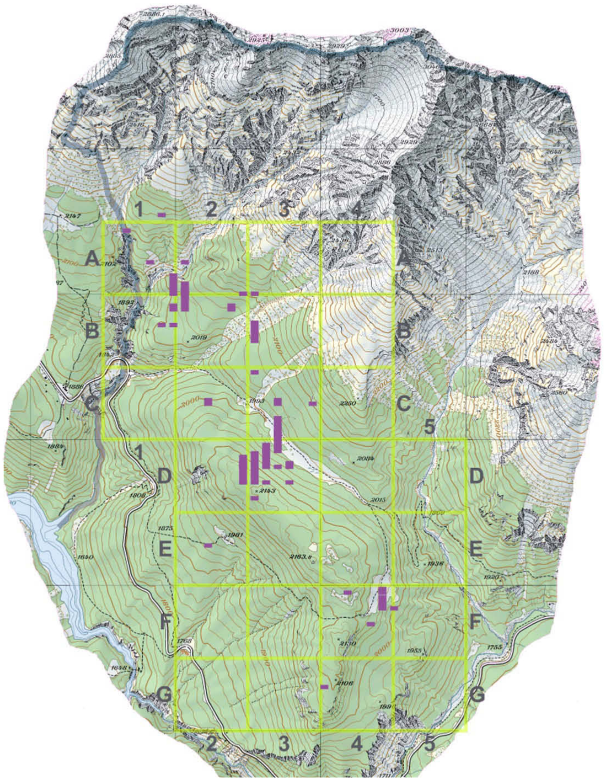


Figure B.1.: Screenshot of setting and data set W1 in 2D

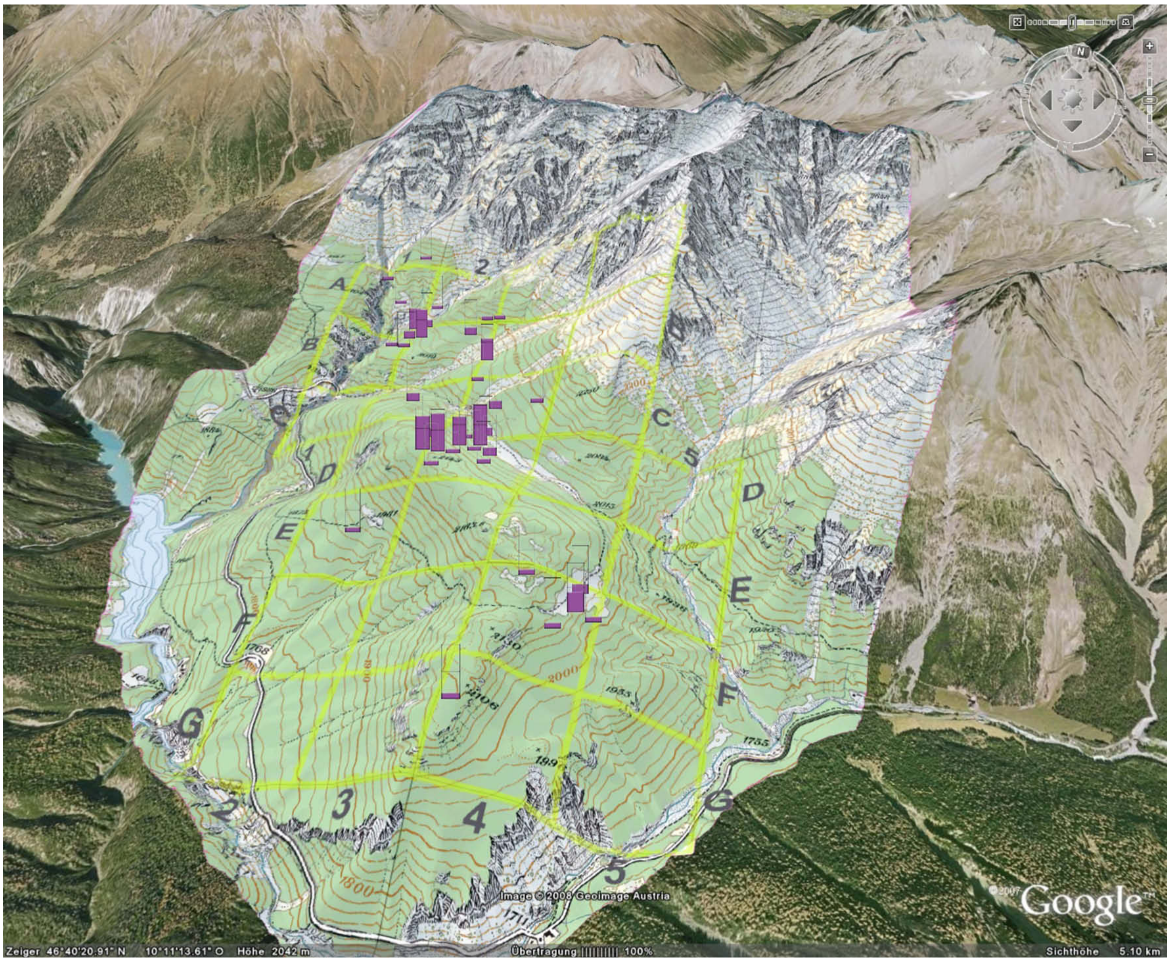


Figure B.2.: Screenshot of setting and data set W1 in 3D

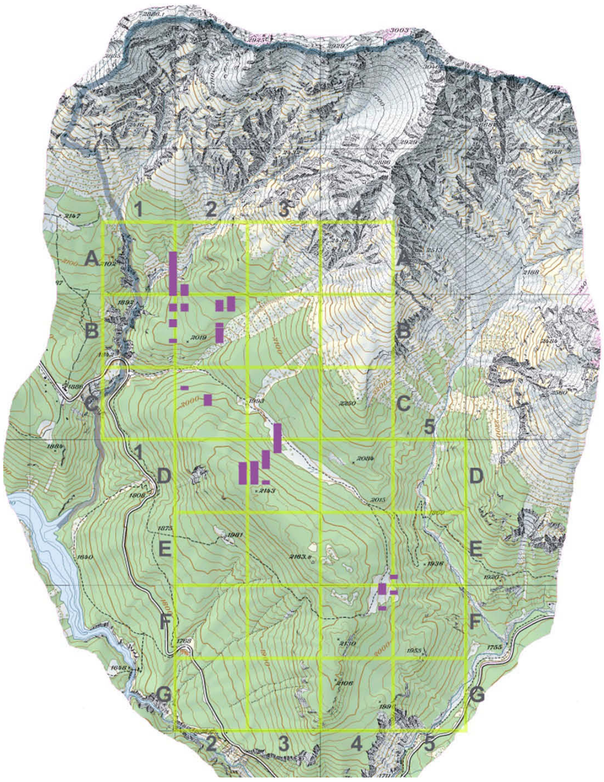


Figure B.3.: Screenshot of setting and data set W2 in 2D

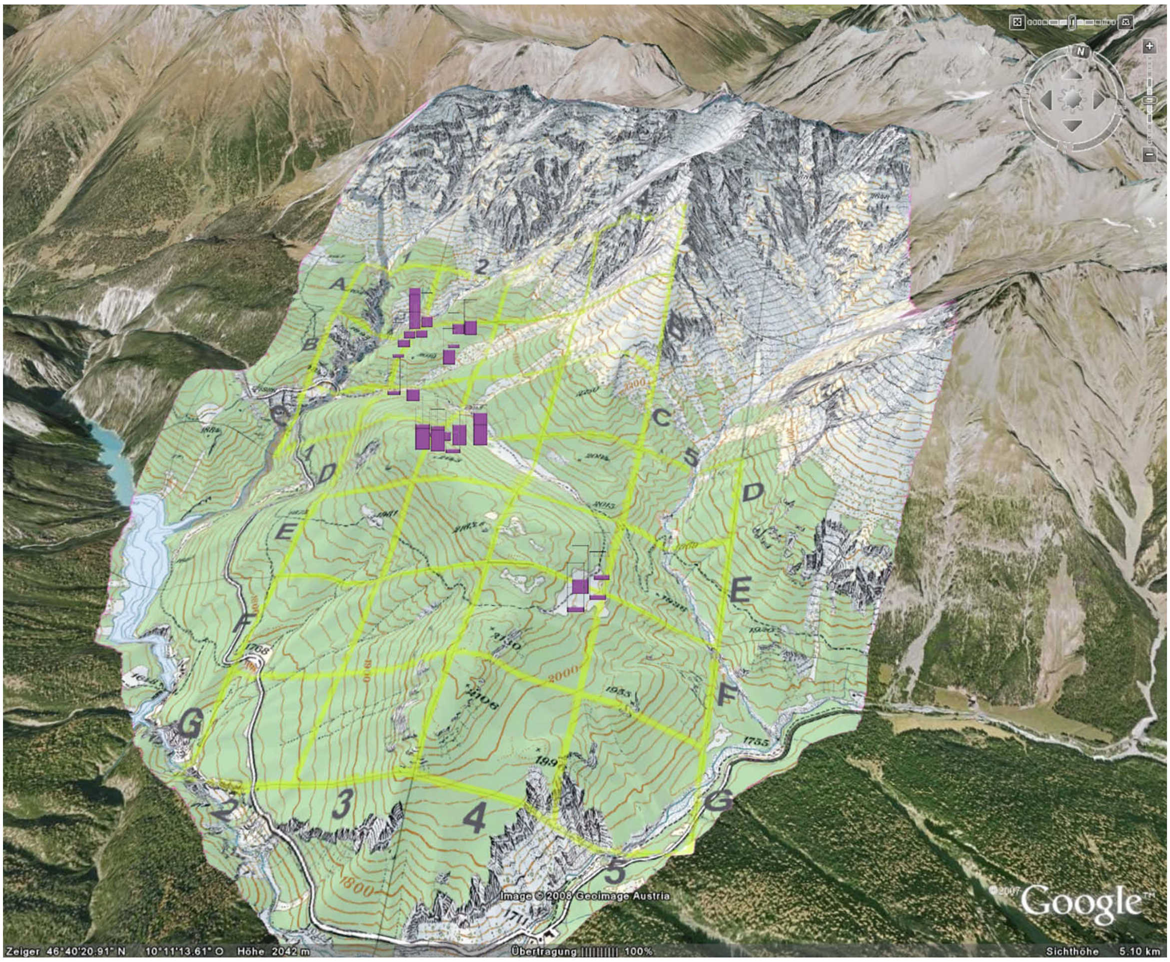


Figure B.4.: Screenshot of setting and data set W2 in 3D

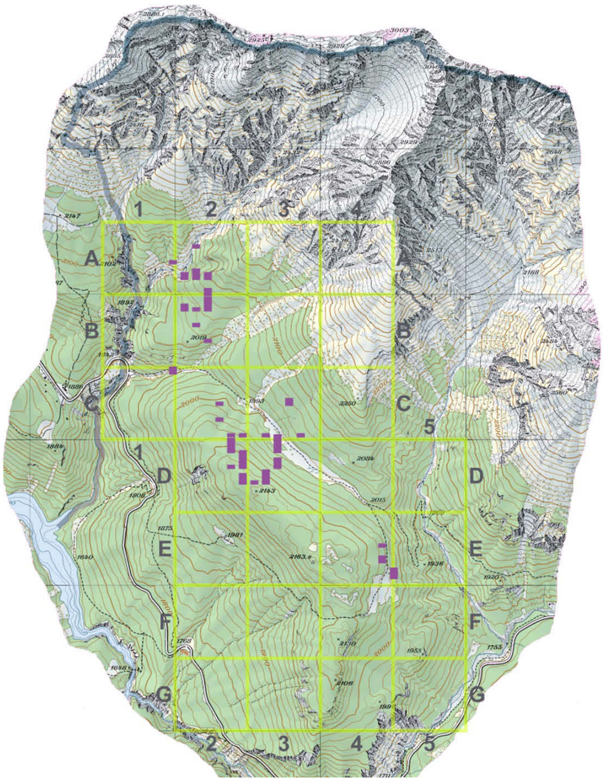


Figure B.5.: Screenshot of setting and data set W3 in 2D

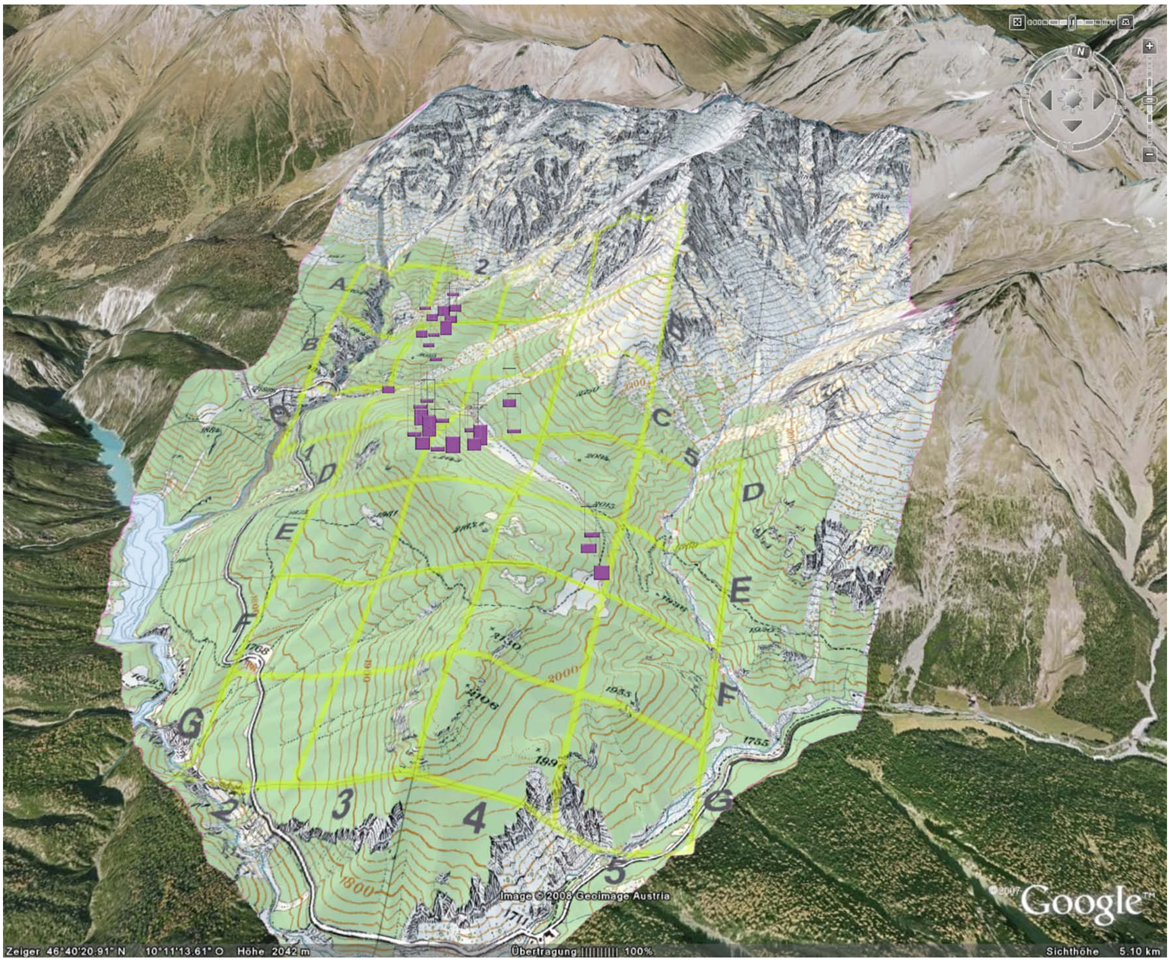


Figure B.6.: Screenshot of setting and data set W3 in 3D

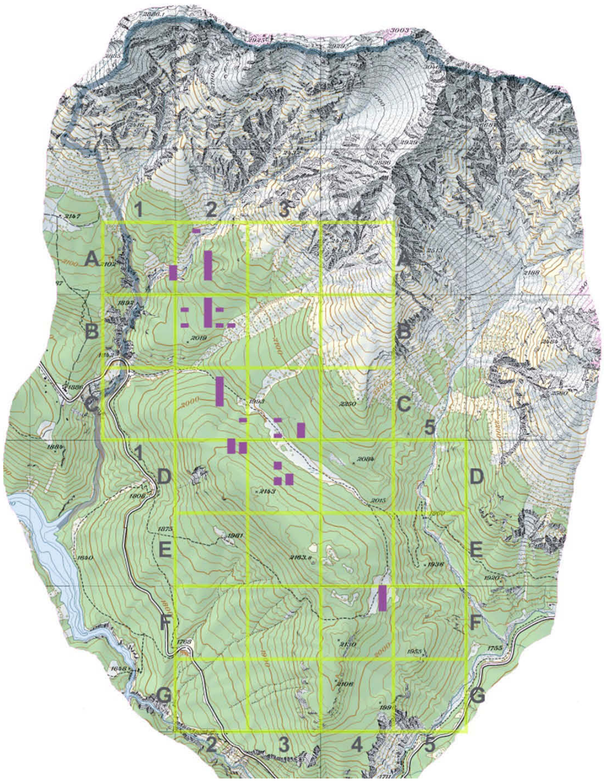


Figure B.7.: Screenshot of setting and data set W4 in 2D

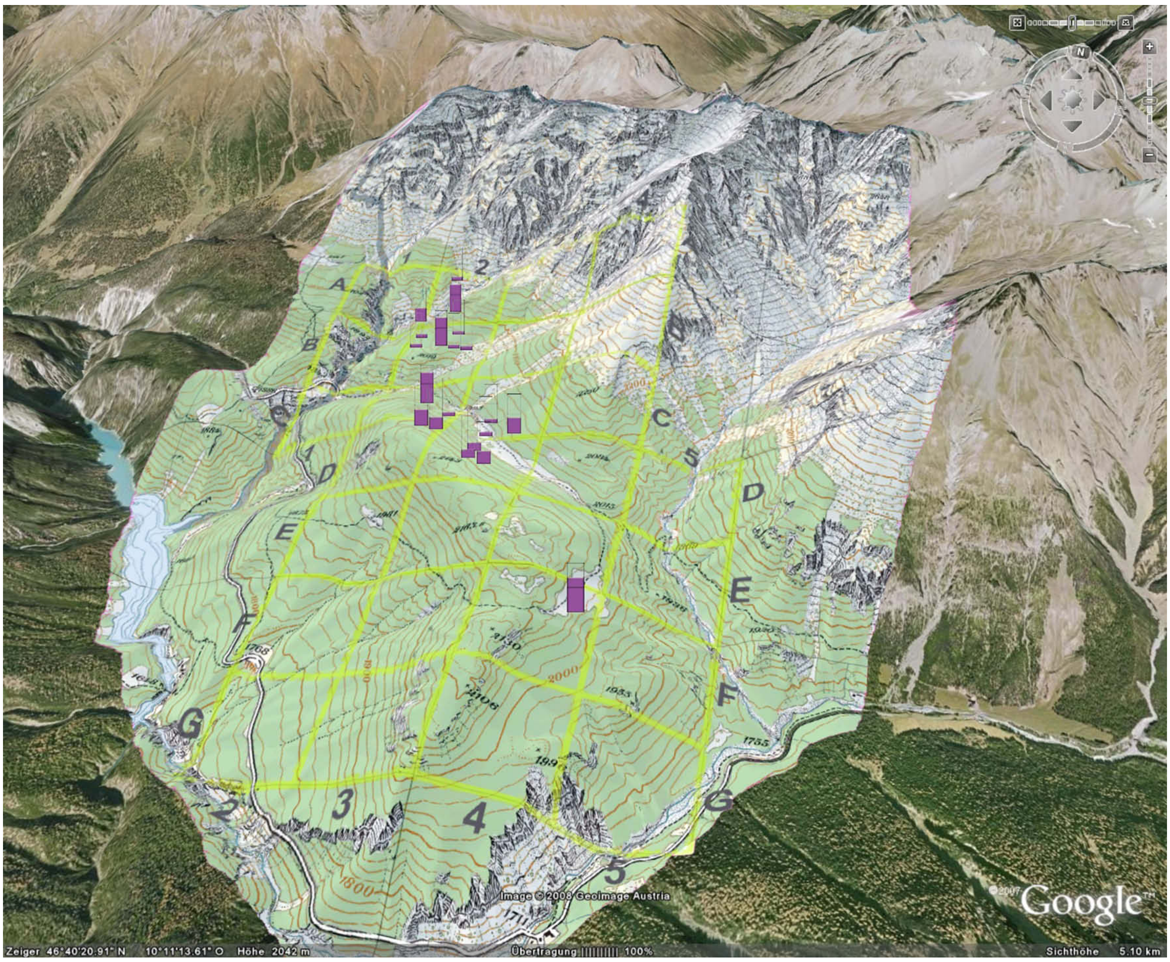


Figure B.8.: Screenshot of setting and data set W4 in 3D

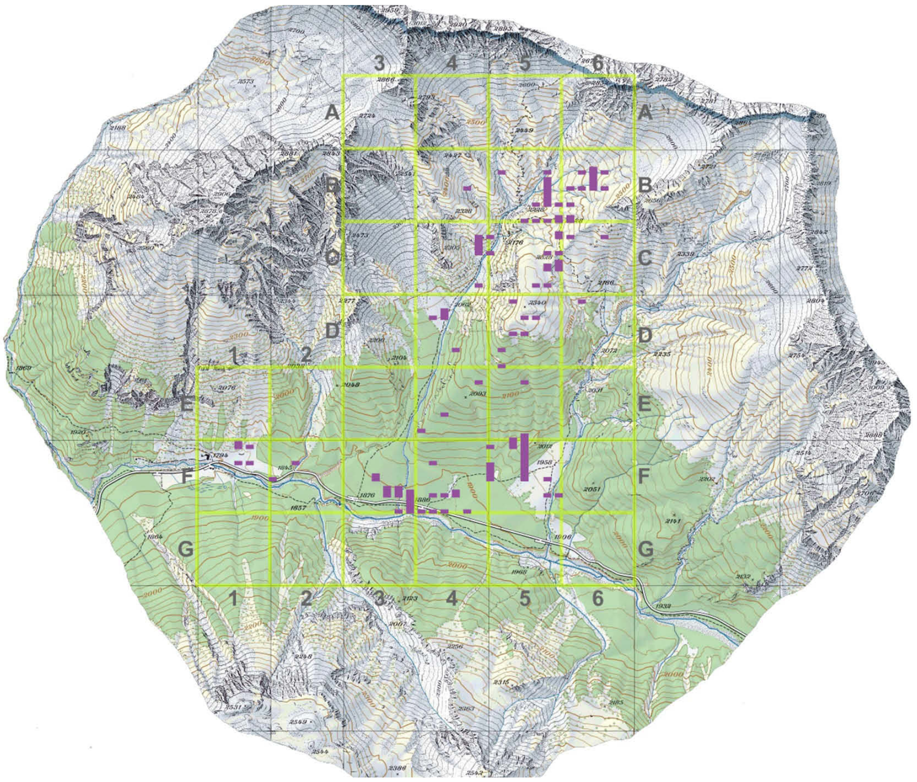


Figure B.9.: Screenshot of setting and data set S1 in 2D

Figure B.10.: Screenshot of setting and data set S1 in 3D

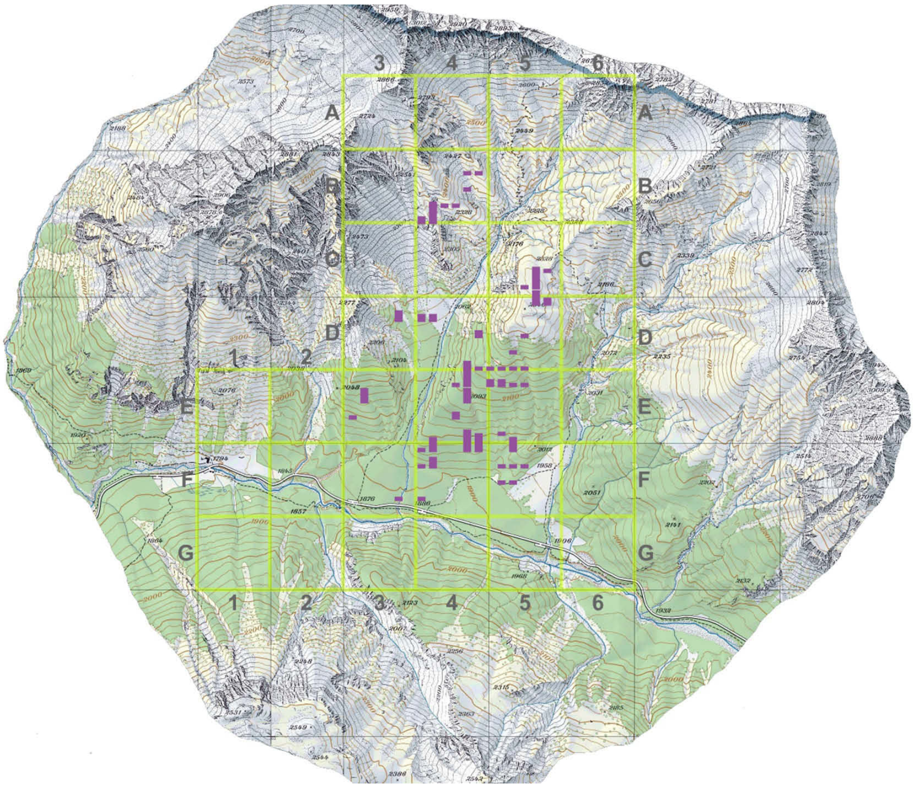


Figure B.11.: Screenshot of setting and data set S2 in 2D

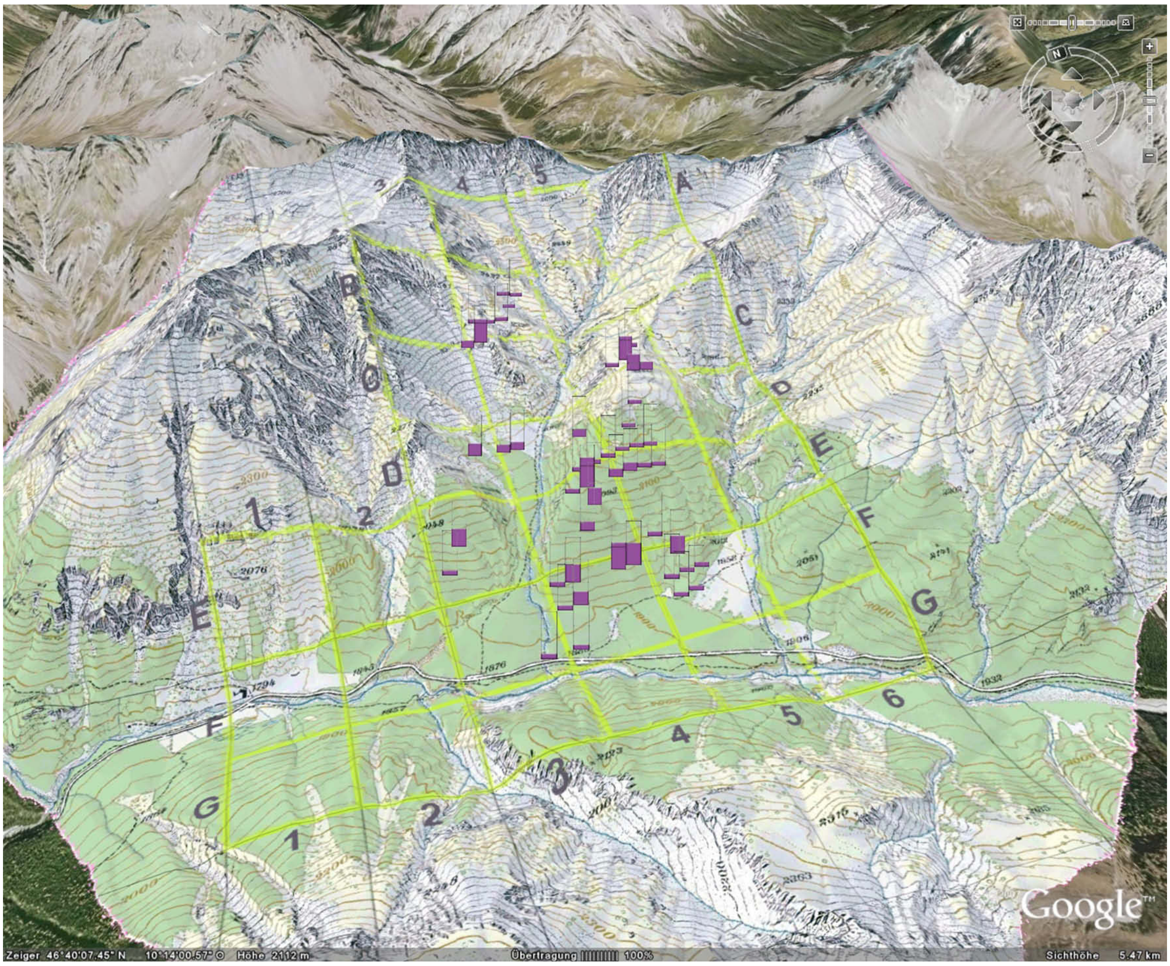


Figure B.12.: Screenshot of setting and data set S2 in 3D

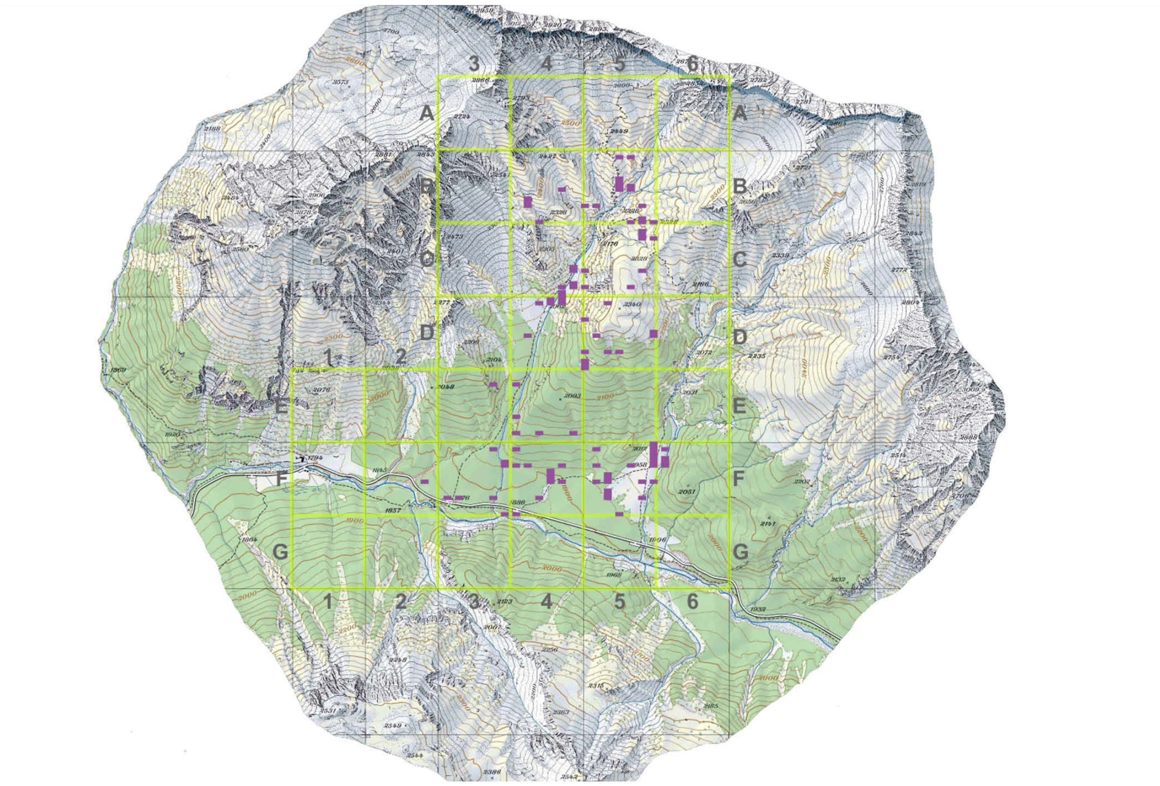


Figure B.13.: Screenshot of setting and data set S3 in 2D

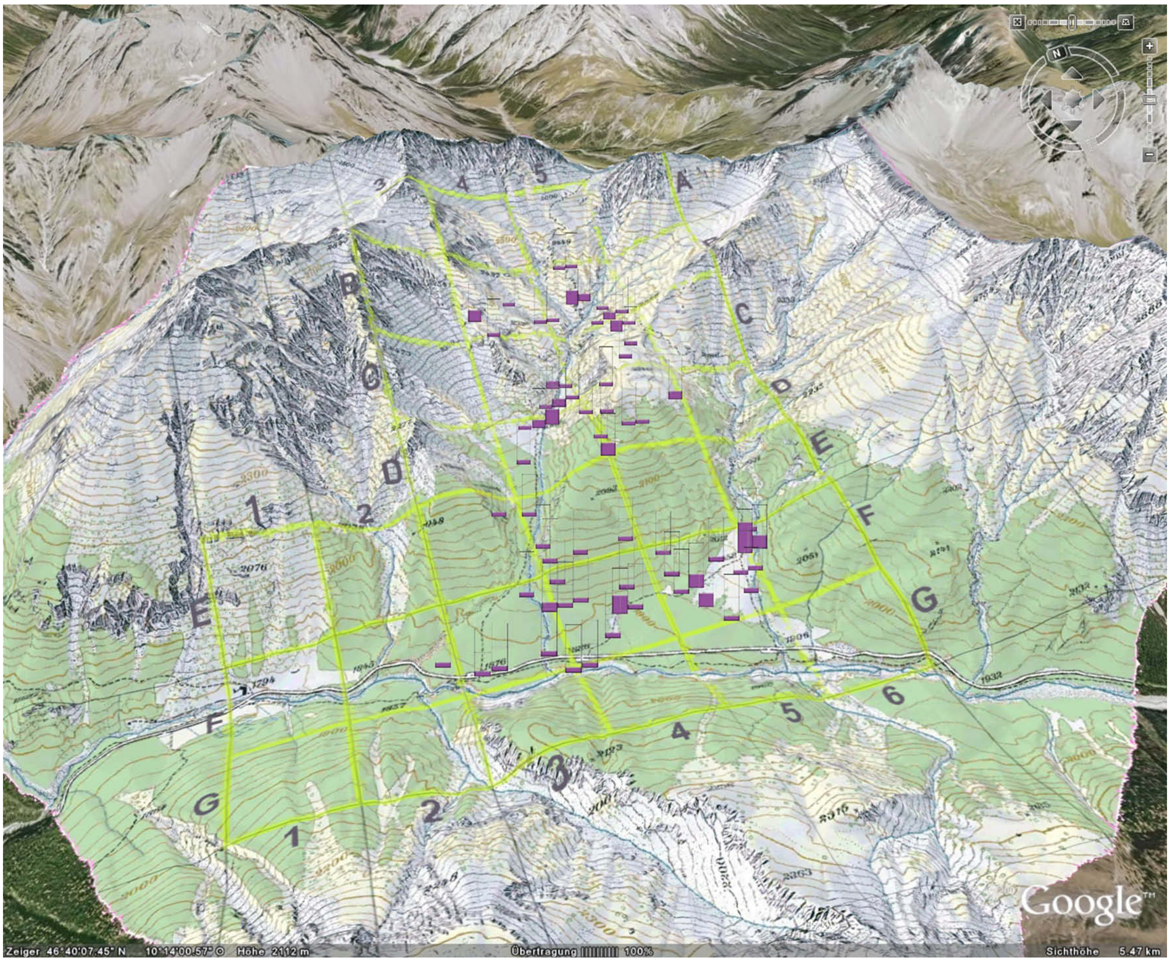


Figure B.14.: Screenshot of setting and data set S3 in 3D

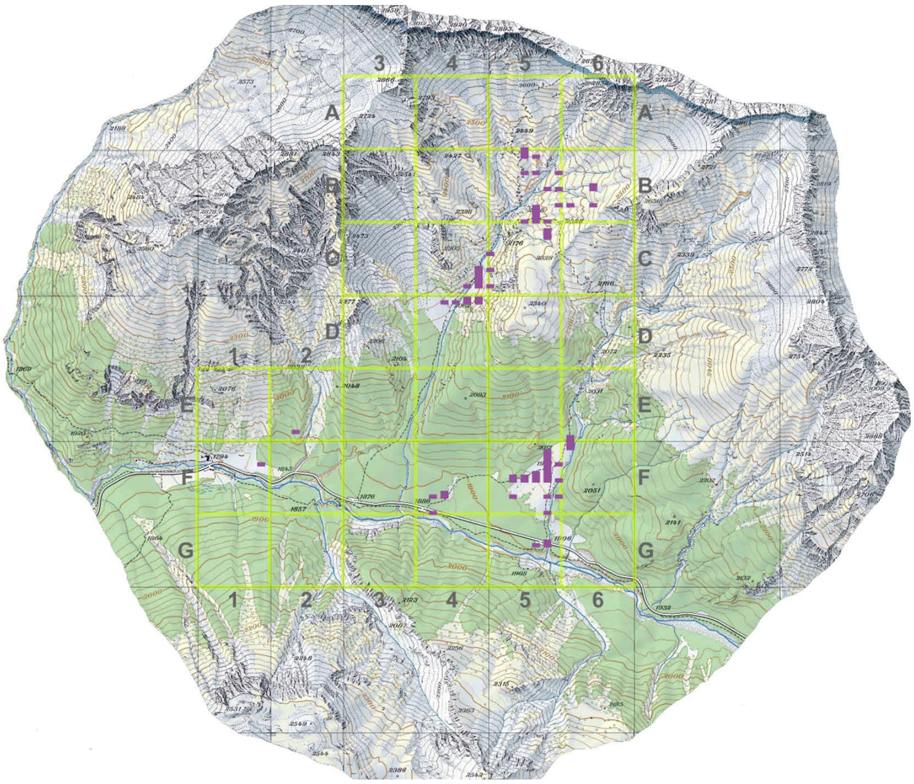


Figure B.15.: Screenshot of setting and data set S4 in 2D

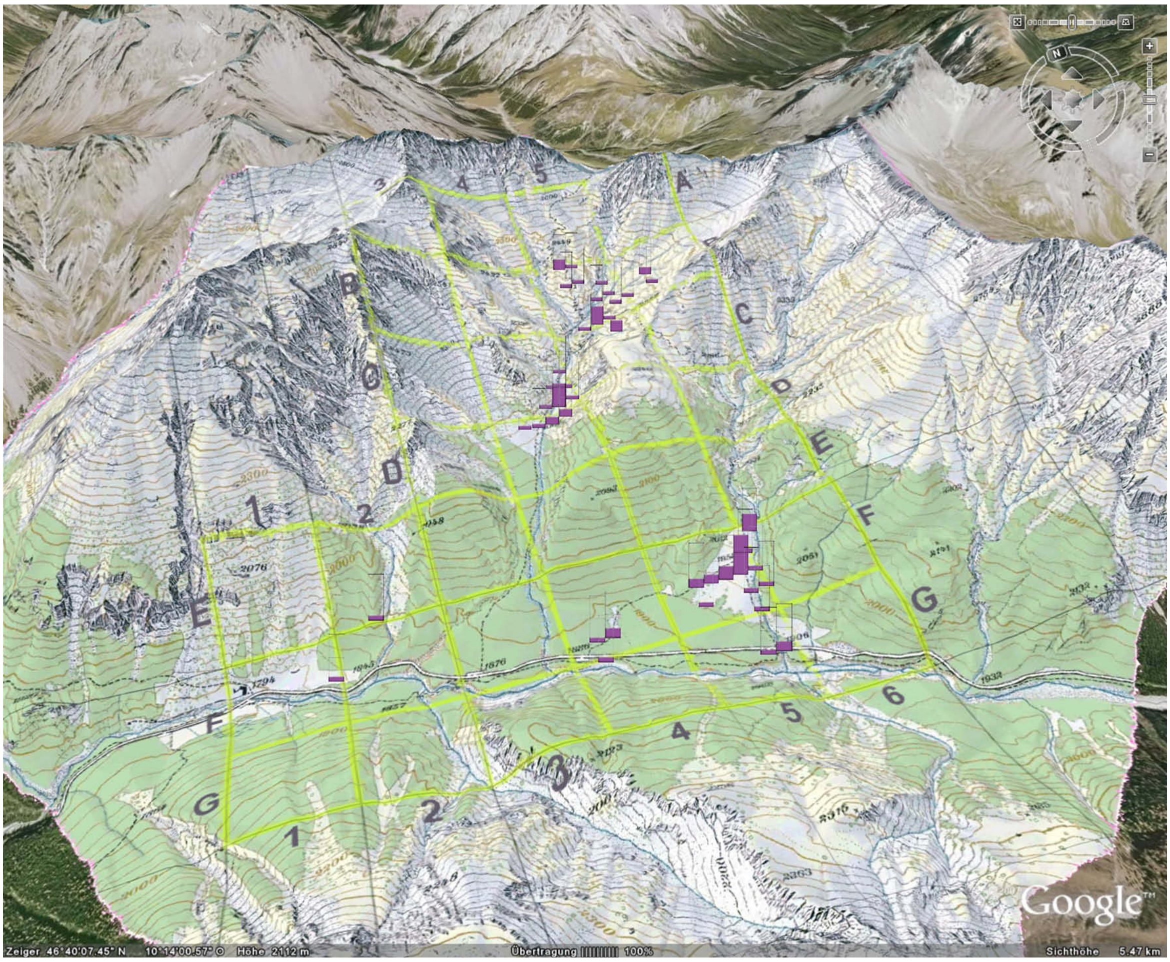


Figure B.16.: Screenshot of setting and data set S4 in 3D

B.2. Stage IIa: characteristics of data sets and settings

		W1	W2	W3	W4
area	total	4.5km^2			
land-cover	forest	80%			
	open (pasture, scree)	20%			
altitude	lowest point	1800 a.s.l.			
	highest point	2500 a.s.l.			
area	data covered	3.75km^2	2.75km^2	2.5km^2	2km^2
altitude	lowest point	1940 a.s.l.	1960 a.s.l.	1900 a.s.l.	2010 a.s.l.
	highest point	2180 a.s.l.	2140 a.s.l.	2140 a.s.l.	2110 a.s.l.
	altitude difference	240m	180m	240m	100m
data values	number	37	23	30	20
	maximum	10	12	5	8
	minimum [quantity]	1 [20]	1 [8]	1 [14]	1 [9]
	distribution of values (values 1 to 13)				
	figures	B.1 & B.2	B.3 & B.4	B.5 & B.6	B.7 & B.8
most appropriate visualisation type (table 3.2)	2D	x	x	x	
	3D				
	2D or 3D				x

Table B.1.: Characteristics of area and data in setting W, data sets W1-W4

		S1	S2	S3	S4
area	total	5.75km^2			
land-cover	forest	55%			
	open (pasture, scree)	45%			
altitude	lowest point	1800 a.s.l.			
	highest point	2800 a.s.l.			
area	data covered	4.5km^2	2.75km^2	4km^2	3km^2
altitude	lowest point	1800 a.s.l.	1890 a.s.l.	1870 a.s.l.	1880 a.s.l.
	highest point	2480 a.s.l.	2360 a.s.l.	2480 a.s.l.	2480 a.s.l.
	altitude difference	680m	470m	610m	600m
data values	number	75	50	68	43
	maximum	13	7	7	9
	minimum [quantity]	1 [55]	1 [30]	1 [51]	1 [29]
	distribution of values (values 1 to 13)				
	figures	B.9 & B.10	B.11 & B.12	B.13 & B.14	B.15 & B.16
most appropriate visualisation type (table 3.2)	2D			x	x
	3D	x			
	2D or 3D		x		

Table B.2.: Characteristics of area and data in setting S, data sets S1-S4

B.3. Stage IIb: data sets and visualisations

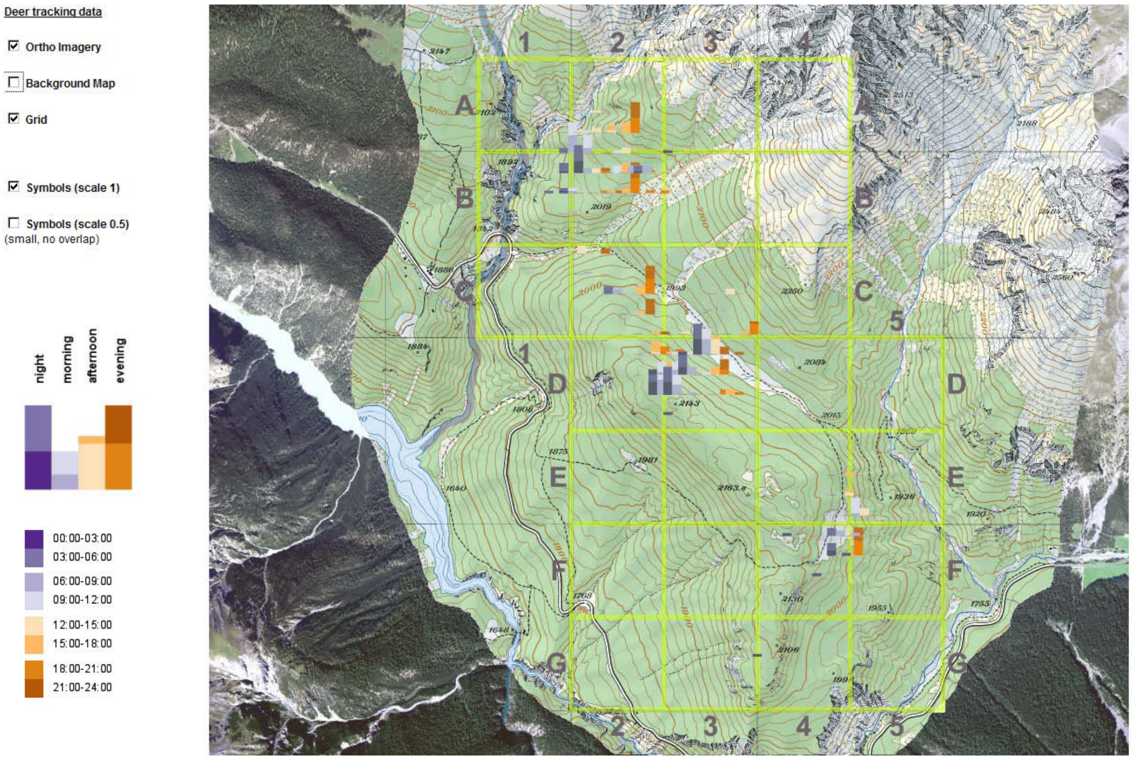


Figure B.17.: Screenshot of setting and data set W in 2D with map background

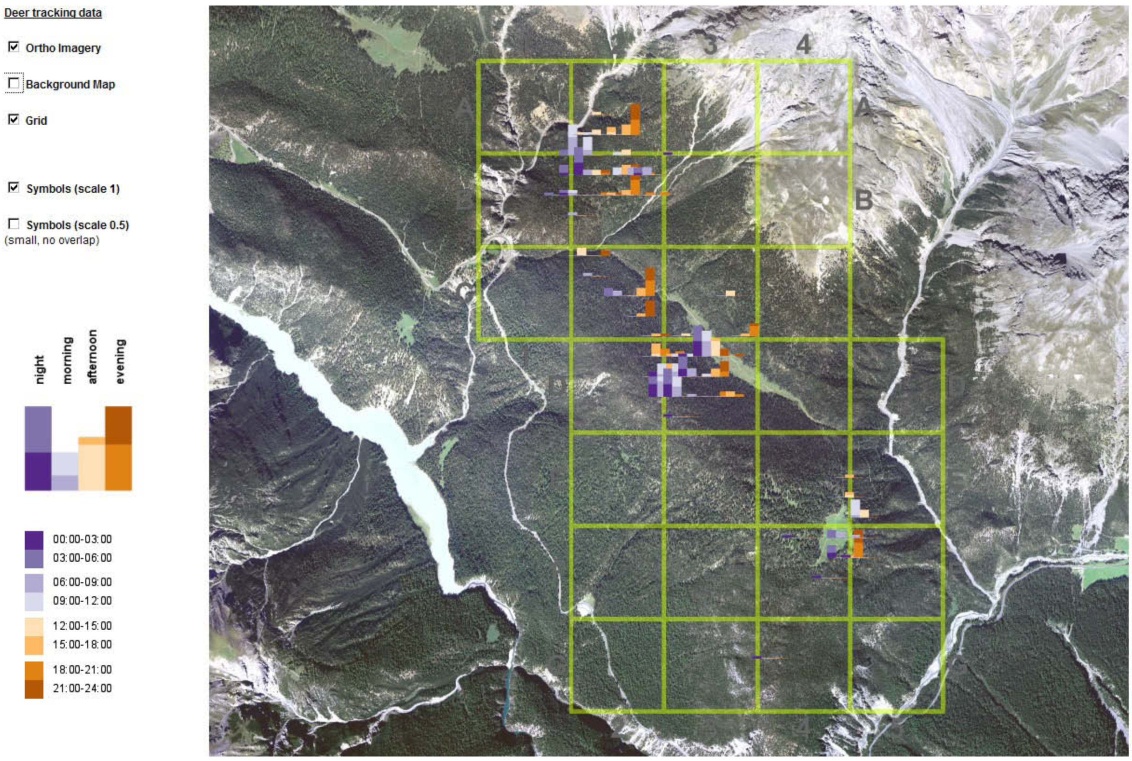


Figure B.18.: Screenshot of setting and data set W in 2D with ortho imagery background

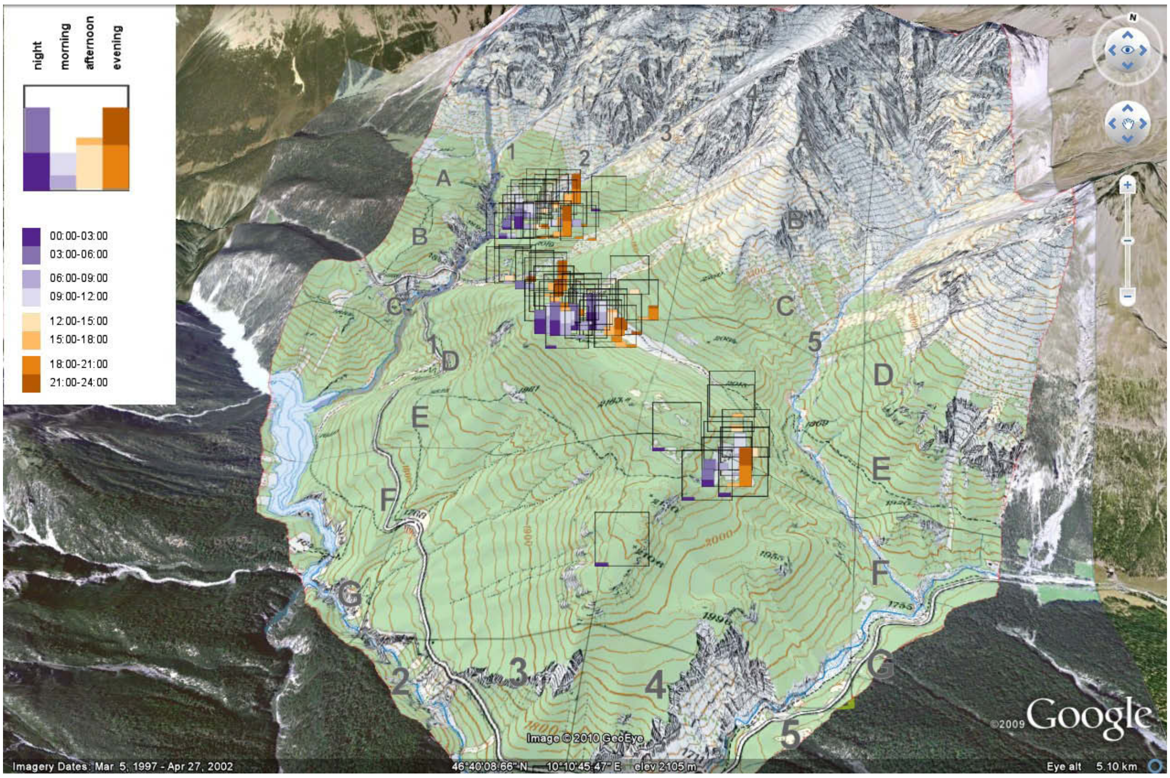


Figure B.19.: Screenshot of setting and data set W in 3D with map background

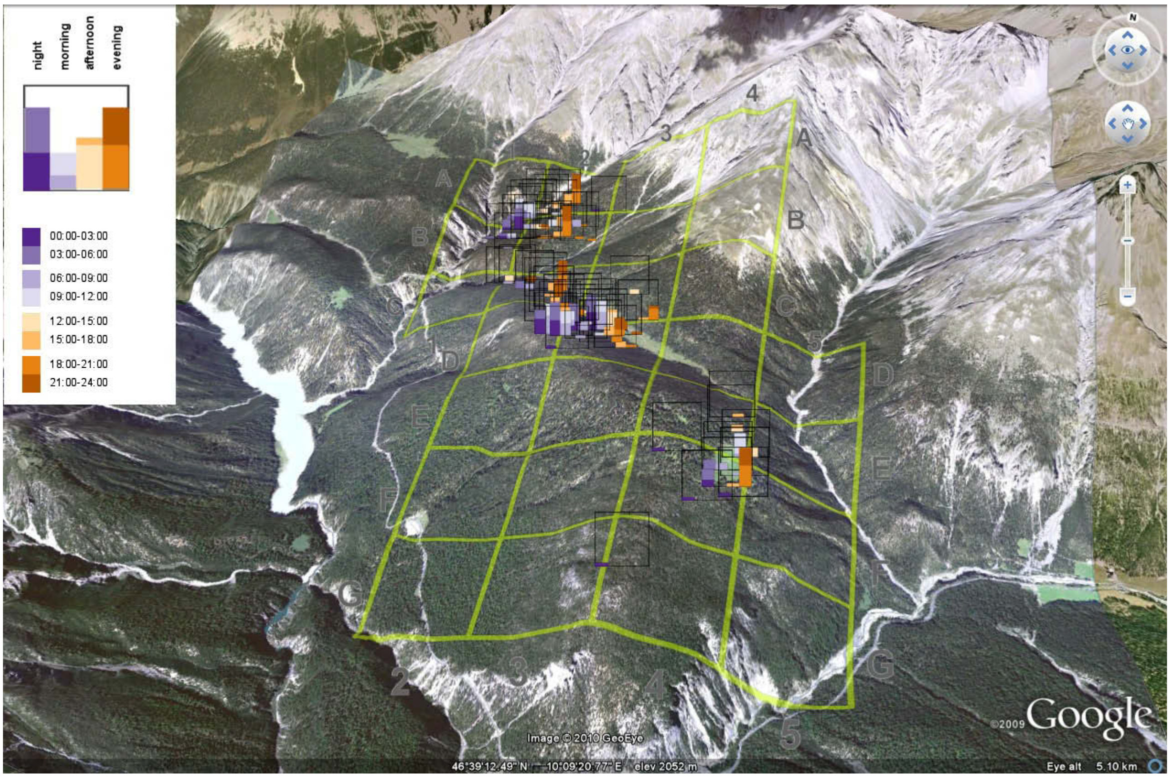


Figure B.20.: Screenshot of setting and data set W in 3D with ortho imagery background

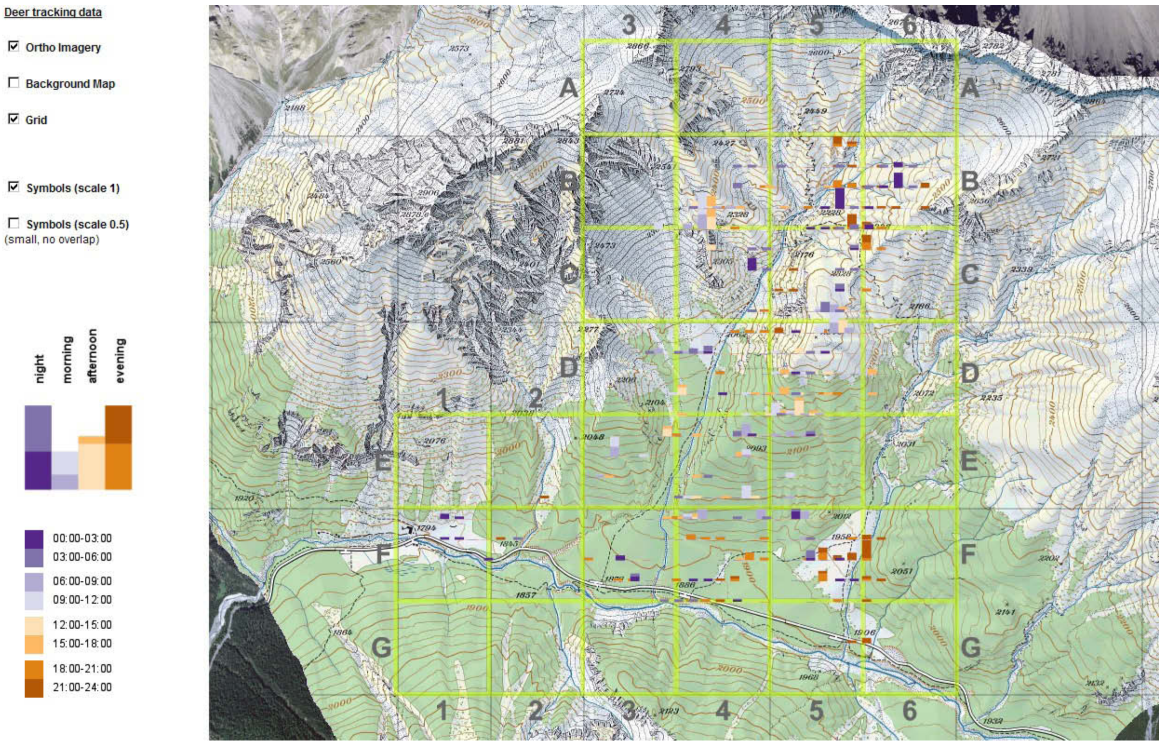


Figure B.21.: Screenshot of setting and data set S in 2D with map background

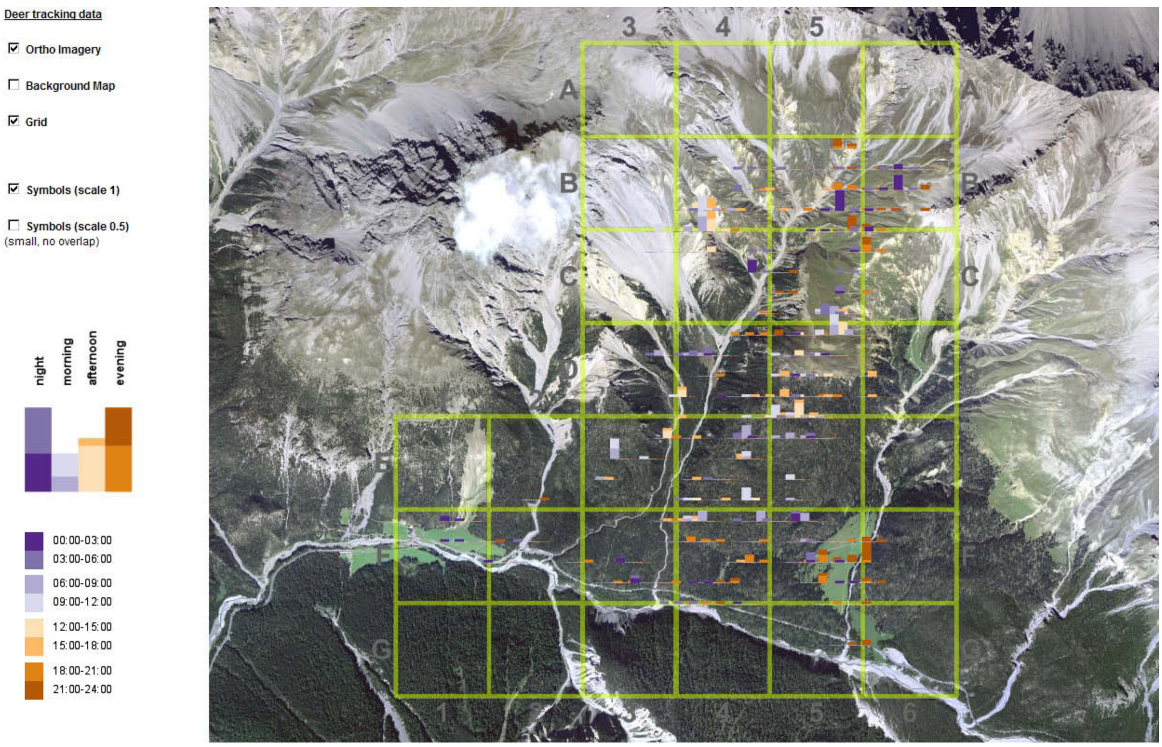


Figure B.22.: Screenshot of setting and data set S in 2D with ortho imagery background

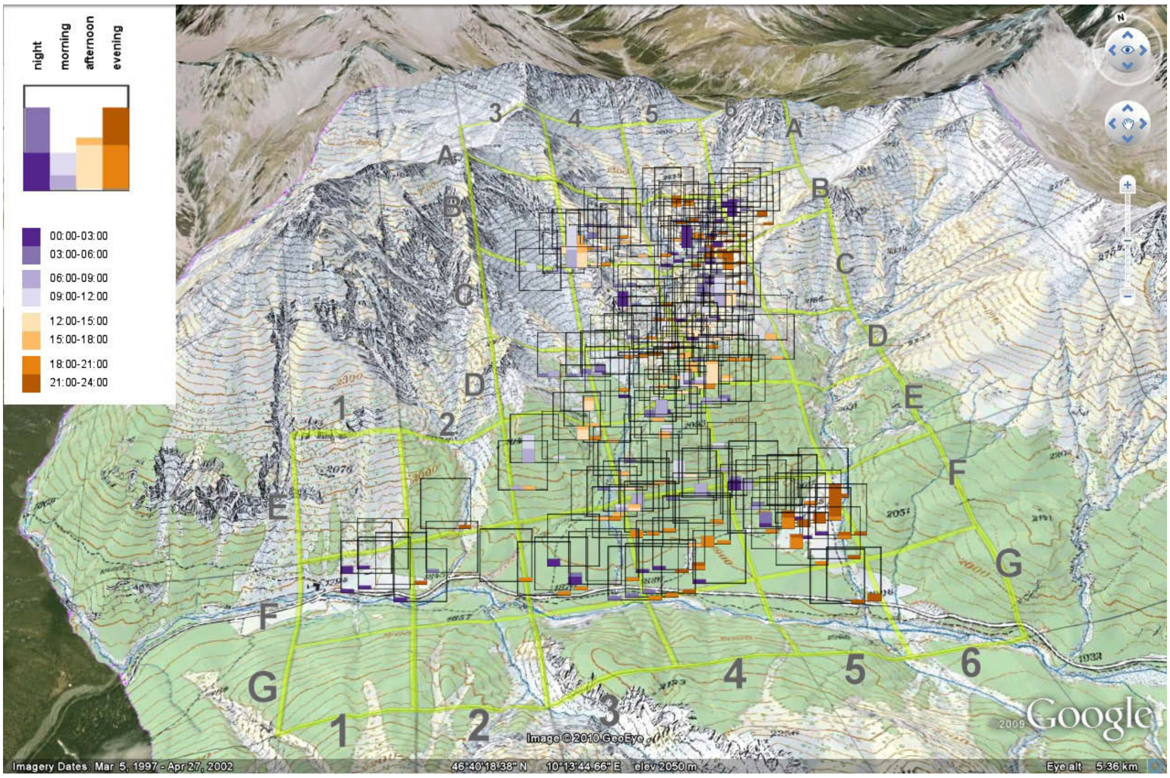


Figure B.23.: Screenshot of setting and data set S in 3D with map background

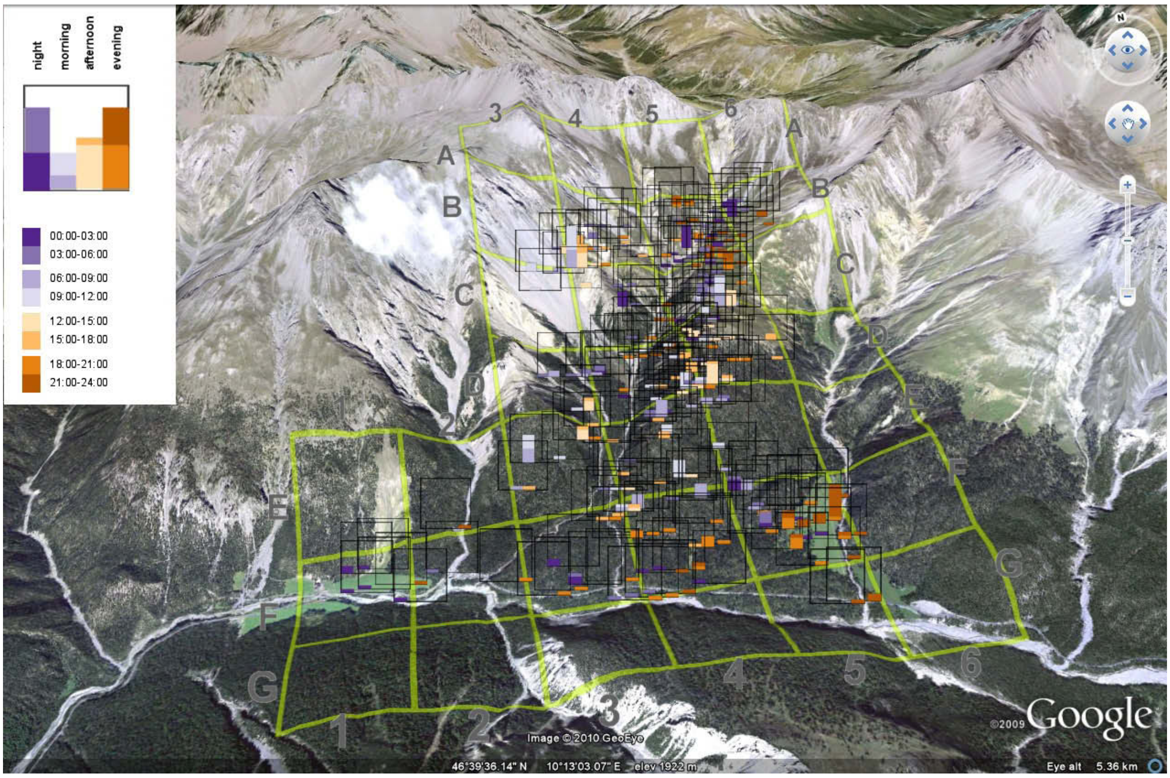


Figure B.24.: Screenshot of setting and data set S in 3D with ortho imagery background

B.4. Stage III: data sets and visualisations

Figure B.25.: Brienz: fs (differences in location) and dh (differences in height) in total

Figure B.26.: Brienz: fs (differences in location) per 5 years

Figure B.27.: Brienz: dh (differences in height) per 5 years

Figure B.28.: Brienz: fs (differences in location) and dh (differences in height) per 5 years

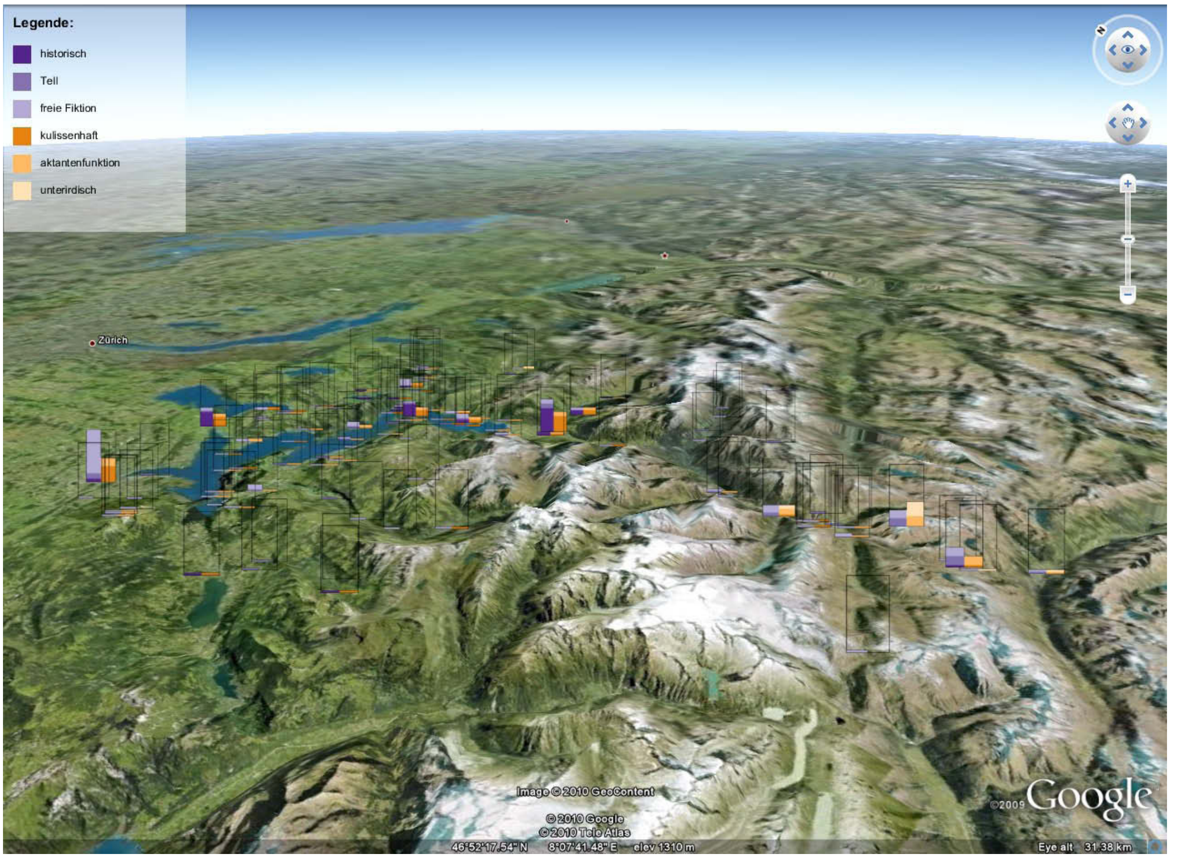


Figure B.29.: Literature Atlas: area Gotthard

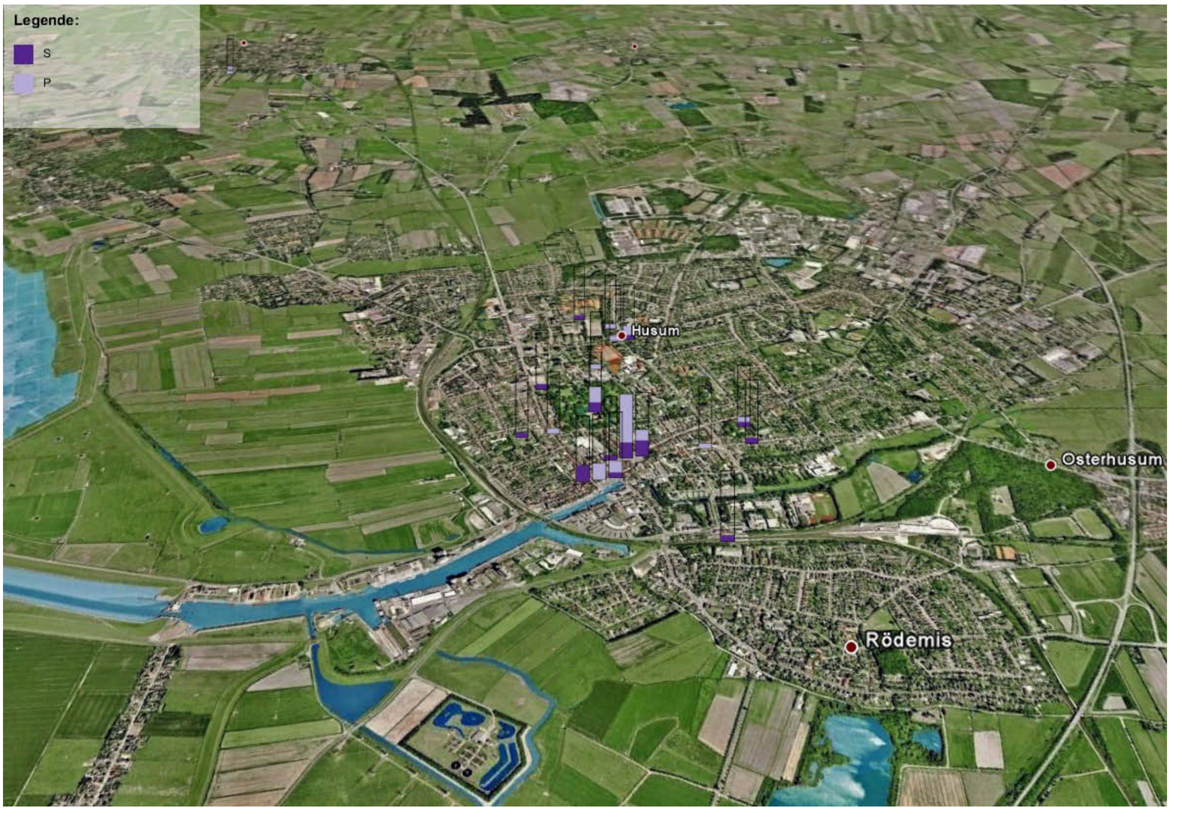


Figure B.30.: Literature Atlas: detail in Husum NFR

Figure B.31.: Literature Atlas: area NFR

Figure B.32.: Literature Atlas: area NFR

Figure B.33.: SNP: Deer 604 in 2004, different seasons

Figure B.34.: SNP: Deer 604 in 2005, different seasons

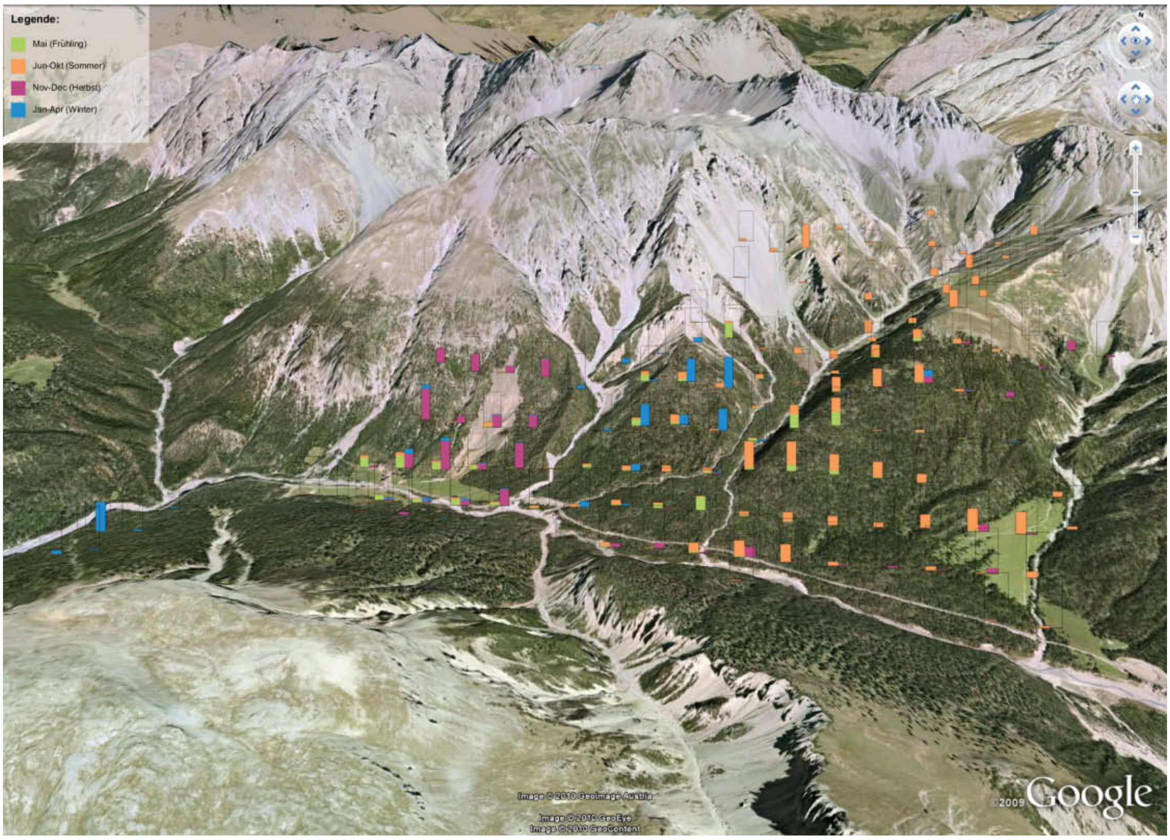


Figure B.35.: SNP: Deer 635 in 2007, different seasons

Figure B.36.: SNP: Deer 635 in winter 2007, different times of the day

Appendix C.

Implementation documents

C.1. Stage IIa

Exemplary invitation to participate in the experiment

Hello fellow former MGI student!
Having completed the MGI course a number of years ago I'm currently working on my PhD at City. I'm evaluating data visualizations in 3D virtual environments with Jason Dykes.
We hope you don't mind me emailing you but I need a favour – and I hope you can help. We're intensively looking for GI professionals and those with experience of GIS who would be willing to participate in an experiment for my PhD study.
The experiment includes judging a number of 2D (SVG) and 3D (Google Earth) visualizations and filling in a questionnaire. It takes place online and will take you about 1 hour to complete.
We hope that participation will be interesting for you – and should get you thinking spatially! The results will help us evaluate the effectiveness of these approaches and contribute to our ongoing Geovisualization research.
I would very much appreciate if you'd kindly consider this request. Please respond to this email and I will send you your personal questionnaire link. Please contact me if you have any further questions.
Thank you,
Susanne and Jason

followed by an email with the personal questionnaire link for people accepting the invitation to participate

Dear xx,
Thank you very much for participating in our study. Your help is very much appreciated. This is your personal questionnaire link [questionnaire link]. Please reserve at least an hour for completing the questionnaire. If possible, make sure you are not disturbed during your work as I'm also recording how long each task takes and you should aim to complete the task successfully as efficiently as possible.
The questionnaire should work with Internet Explorer and Firefox. However, some participants have reported problems with the display of SVG in Firefox. You may want to try it with Firefox or use Internet Explorer from the beginning.
Please note that the data collected in the questionnaire will be used anonymously only. It would be great if you could fill in the questionnaire before 30 November 2008 so that I can continue with my research to plan. Please email me informally if, for any reason, you will not be able to participate.
Thanks again,
Susanne

Balancing questionnaire assignments

The participants who responded positively to the invitation were assigned in a balanced way to the different questionnaires but some of them did not participate (mainly because of lack of time, some software problems) despite their original acceptance of the invitation to participate.

A1 QuestA1	A2 QuestA2	A3 QuestA3	A4 QuestA4	B1 QuestB1	B2 QuestB2	B3 QuestB3	B4 QuestB4
think-aloud	think-aloud		think-aloud		think-aloud	think-aloud	
✓✓✓	✓✓✓✓✓	✓✓✓✓✓	✓✓	✓✓✓✓✓	✓✓	✓✓✓✓	✓✓✓
✗✗✗✗✗	✗✗	✗✗✗✗	✗✗✗✗✗	✗✗✗✗✗	✗✗✗✗✗✗	✗✗✗	✗✗✗✗✗
1+3+5=10	1+5+2=9	0+5+4=9	1+2+5=9	0+5+5=10	1+2+6=10	1+4+3=9	0+3+5=8

in total 74 invited participants (✓ = participated, ✗ = did not participate)

Questionnaire introduction

QuestA1

Javascript-Test - if there is no date/time visible below, please change your browser settings to allow Javascript and then reload this page. Thu Mar 12 2008 12:55:38 GMT+0100

This study evaluates the possibilities of 2D and 3D visualisations for the visual exploration of data sets. For this, we visualise deer tracking data in different areas and ask a number of questions. It is your task to answer those questions as accurate and detailed as reasonable. The number of deer visits at a specific location are visualised through the use of bars coloured violet (the taller a bar is the more often the deer has visited the specific location). We record how long you take to answer each of the questions. But please take as much time as you need for answering the questions. However, we'd like to ask you to concentrate on the questions and not to wander off while working through the questionnaire. Answering all questions will take you about one hour.

The questions are mostly related to either an area or an altitude. Area means something bigger than a single location or point in the landscape. Altitude means a specific altitude (e.g. 2410 metres above sea level) or an altitude range (e.g. 2000-2200 metres above sea level). It is your choice how you will describe the areas and altitudes in your answers. However, you may want to use the information from the background map or the overlaid grid for your descriptions.

Thank you very much for taking part in this study!

Remark: The following button navigates you from page to page in this questionnaire. It is not possible to navigate backwards. Using the 'Back' button of the Browsers causes the questionnaire to be reloaded and all already given answers will be lost.

Exemplary structure of questionnaires

questionnaire introduction followed by a page with an exemplary 2D SVG visualisation and instructions on how to interact with it

First a 2D example to introduce you to the visualisation technology. You need the SVG viewer to view the representation. [image 'Hallo']
If you cannot see the 'Hallo' above you will need to download and install the SVG viewer version 3.03 from http://www.adobe.com/svg/viewer/install/ (scroll down to the downloads).

Please open the following SVG visualisation: 2Dw0.svg (leave it open, you will need it for answering the subsequent questions) - if you have a slow internet connection it may take some time until the background map is loaded. The visualisation should look like the following screenshot. If not - please install the SVG viewer again, try another Browser or write an email to [email address] reporting your problems.

[screenshot of the SVG visualisation 2Dw0.svg]

Navigation:
- pan the visualisation by pressing the 'Alt' key and dragging the pressed left mouse key in the visualisation
- zoom in by pressing the 'Ctrl' key and clicking into the visualisation
- zoom out by pressing the 'Ctrl' & 'Shift' keys and clicking into the visualisation

followed by exemplary questions and confidence ratings ("Describe the location of the bar.", "How do you rate your confidence in this answer?" and "Describe the altitude of the bar.", "How do you rate your confidence in this answer?") to train for the tasks to be answered during the experiment

followed by a page with an exemplary 3D SVG visualisation and instructions on how to interact with it

Please close the 2D visualisation.

Now an example to introduce you to the 3D visualisation technology. You need Google Earth installed on your computer to view the visualisation. If Google Earth is not installed on your computer you can download it from http://earth.google.com/ (free version) and install it. If Google Earth is already installed on your computer then please switch of all content and presentation layers except the terrain layer (remove all ticks, see the illustration below).

Please open the following 3D visualisation: 3Ds0.kml (leave it open, you will need it for answering the subsequent questions) - please be patient, it might take some time to start up Google Earth and load the visualisation completely

The visualisation should look like the following screenshot. If not - please install Google Earth again, switch of all layers (remove all ticks in the left hand side menu except the terrain layer) or write an email to [email address] reporting your problems.
[screenshot of Google earth with 3Ds0.kml opened]
Please familiarise yourself with the navigation. Especially, try to navigate and rotate the view in the oblique view.
Navigation:
- drag the 'N' with the left mouse key to the left or right to rotate the view around the current center of the view
- drag the mouse while pressing the 'Ctrl' key to change the angle of the oblique view or alternatively, use the '+' and '-' buttons of the slider

followed by exemplary questions and confidence ratings ("Describe the location of the bar.", "How do you rate your confidence in this answer?" and "Describe the altitude of the bar.", "How do you rate your confidence in this answer?") to train for the tasks to be answered during the experiment

followed by several pages asking to open a visualisation (2D or 3D), answering the tasks and rating the confidence in the answer as done for the exemplary 2D and 3D visualisations

C.2. Stage IIb

Questionnaire introduction (in German)

Set 1

benötigte Software: SVG Viewer und Google Earth

Experiment im Rahmen meiner Doktorarbeit. Es soll die Nützlichkeit und Verwendbarkeit von virtuellen 3D-Umgebungen (Google Earth) für die explorative Informationsvisualisierung im Vergleich zu 2D (interaktives SVG) evaluiert werden. Für die Aufgaben bitte nacheinander folgende zwei Fragebogen ausfüllen. Bitte die Instruktionen genau lesen und befolgen.
Die dargestellten Daten sind Hirschtrackingdaten aus dem Schweizerischen Nationalpark. Ein Hirsch wurde im Sommer (drei Monate) und im Winter (drei Monate) getrackt und die Daten dargestellt (Balkenhöhe = Häufigkeit der Aufenthalte an einem Ort). Die Aufenthalte sind farbcodiert bezüglich der Tageszeit.
Als Hintergrund kann entweder die Karte 1:25'000 oder das Orthophoto verwendet werden. Das Grid dient als Hilfe für die Beschreibung der Orte und Räume von Erkenntnissen aus den Daten.

Sommer-Daten: [Link zu So3D]
Winter-Daten: [Link zu Wi2D]

Notes for verbal instructions on experiment and software interaction (in German)

Frage: Analysiere die Hirschdaten und beschreibe deine Erkenntnisse in Bezug auf Lage, Höhe und Tageszeit.

Verwende einen neuen Eintrag pro Antwort und bewerte jeweils wie sicher du in Bezug auf diese Antwort bist.
- nur Vorwärtsnavigation möglich (nicht den Back-Button des Browsers verwenden)
- 30 Minuten Zeit

danach das zweite Setting bearbeiten (2D oder 3D respektive)

Navigation Google Earth:
- zuerst alle Layers ausschalten (ausser Terrain)
- kurz ausprobieren, z.B. mit dem Navigator rechts oben
- Zieh mit gedrückter linker Maustaste das 'N' nach links oder rechts um die Ansicht um das momentane Bildzentrum zu rotieren
- mit gedrückter Ctrl-Taste und gedrückter und gezogener linker Maustaste in Google Earth, kannst du die Steilheit der Schrägansicht variieren - oder alternativ
- den Slider oder die '+' und '-' Buttons benützen
- Möglichkeit Hintergrundinfos ein-/auszuschalten

Navigation SVG (im IE mit SVG player):
- Ctrl + Maus ⇒ Zoom in; Ctrl + Shift + Maus ⇒ Zoom out
- Alt + Maus ⇒ Pan
- rechter Mausclick ⇒ diverse Optionen
- links: Möglichkeit Hintergrundinfos ein-/auszuschalten und Symbolgrösse verändern (für Detailanalyse, ohne Overlap)

C.3. Stage III

Invitation to participate

Contacts to potential data experts for different case studies was sought early in the research process. After selecting the divers cases the data experts were contacted again for the final agreement of participation in the research. Following an exemplary email (in German).

> Liebe xxx
> [. . .] Nun komme ich nach einigen Verzögerungen endlich zur letzten Phase meines Projekts und somit zu den 'real-world'-Fallstudien mit Datenexperten. Dafür würde ich gerne nochmals die Hirschtracking-Daten visualisieren und dann die Visualisierungen mit dir besprechen. Kann ich dich für eine Mitarbeit in dieser Fallstudie gewinnen?
> Nachstehend eine Zusammenstellung welche Aufgaben auf dich zukommen würden und der ungefähre Aufwand.
>
> Teil 1:
> Beschreibung des Datensatzes – ein Word-Dokument ausfüllen (etwa 45min)
> Online Umfrage allgemein zu 3D-Visualisierungen (etwa 20min)
>
> Teil 2:
> Ich erstelle verschiedene 3D-Visualisierungen deiner Daten (kein Aufwand deinerseits)
>
> Teil 3: Ich besuche dich und wir besprechen die 3D-Visualisierungen (1-2 Std)
> Debriefing etwa 1 Woche später per Email oder Telefon (etwa 30min)
>
> Herzlichen Dank für deine (positive :-)) Antwort und liebe Grüsse
> Susanne

Case study protocol

The different phases of the case study are labelled in italic font in the following internal case study protocol. Instructions or reasons are given in square brackets [right aligned].

> *phase I – re-contact experts of selected cases*
>
> [personal or by phone depending on distance]
>
> □ determine case study schedule (fix appointments etc.)
> □ ask for current data set to be sent (if not already available) or specific/interesting sub data set
> □ provide access to experiment I (3D with frames part of stage I only)
>
> [baseline participants]
>
> in the same or a second contact (according to experts preferences)
> □ ask the data questions (see data collection shells for the exact questions)
> *phase II – create 3D visualisation*
>
> [use the data sets and potentially input from the questionnaire
> to create the 3D visualisations (design similar to stage IIa or stage IIb)]
>
> *phase III – visit the experts in their work environment*
> □ set up 3D visualisation with them (provide technical support,. . .)
>
> [gather first impressions/reactions]
>
> □ let them play with the visualisation and ask the questions (see data collection shells for questions)
> □ they fill in a short suitability questionnaire
> □ thank them for participating in the study

phase IV – final contact
□ leave them with the 3D visualisation (for informal use, no further tasks set)
□ contact them again after 1 week (see data collection shells for questions, asked per email)

[capture after-thoughts]

□ thank them again for taking the time to participate

Data collection shells

case questions: phase I (descriptive)

description of data set and context	What data does the data set contain? [description] How was the data set collected? Why was it collected? What are the most important aspects of the data set? When/in what situation is the data related to landform? Is this important? Why?	*data*
2D / current visualisations	What kind of visualisations of the data set do you currently use (if any)? If not, why don't you use visualisations of your data? Are the visualisations available for this study?	*data*
previous/current analysis of the data	What is your interest in the data set? How do you currently explore and analyse the data set? What are (typical) questions you want to answer? What are (typical) tasks? Are there any (formal or informal) reports or writings of the findings? Are they available for this study?	*data*
current problems	What problems do you currently face when exploring/analysing the data set? [with regards to analysis, tasks, visualisations] Are there any research questions you cannot answer?	*data*

case questions: phase III

case & 3D vis	Explain/show me your data: - What are important patterns? - How and where is your data related to the landscape? (think about the 'triangle' attribute – space - topography)	*data*
	Can you answer your research questions? What are the answers? Do you want to ask new research questions? Do you see new opportunities for data analysis?	*data*
	Any new insights?	*data*
comments		*data*

suitability questions (rated on a seven-point scale from 'very suitable/very much' to 'not suitable/not at all')

How suitable is this type of visualisation for the exploration of your data?
Do you like the visualisation?
Would you like to use this type of visualisation for other similar data sets?

case questions: phase IV

debriefing	Did you use/play with the 3D visualisation during this week?	*data*
	Did you find any new insights into your data?	*data*
	Do you have any further comments or questions regarding the 3D visualisation or the study in general?	*data*