COMMUNICATING PERCEIVED GEOSPATIAL DATA QUALITY OF 3D OBJECTS IN VIRTUAL GLOBES

KRISTA JONES

COMMUNICATING PERCEIVED GEOSPATIAL DATA QUALITY OF 3D OBJECTS IN VIRTUAL GLOBES

by

© Krista Jones

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Science

Department of Geography

Memorial University of Newfoundland

September 2011

St. John's Newfoundland

Abstract

The emergence of Web 2.0 and applications such as on-line mapping and virtual globes (VGs) have significantly changed the way geospatial data are used. While the use, creation and analysis of geospatial data was once the realm of experts, these recent technological advancements have led to the democratization of geospatial data. Maps can now be produced by the contribution of thousands of users. This creates challenges when one wants to rely on such data to make a decision, as the quality of the geospatial dataset can be highly heterogeneous or even unknown. This thesis presents an approach for visualizing the perceived quality of 3D objects in VGs. The visualization approach developed adapts an approach successfully used in other domains, the five star and user feedback system, to communicate the reliability or quality of data/products to lay users. This approach helped identify a method of communication which is both technically feasible and easily understandable by both expert and non-expert users alike, while not oversimplifying the quality information being communicated. Determining the appropriate ways in which to visualize this approach in a 3D geospatial environment was then done through implementing prototypes and performing a user evaluation. The results of the user evaluation suggest that the preferred visualization method uses a number within a star to show the object's quality, while the least preferred method is the typically used visualization using five individual stars.

Acknowledgments

I would like to thank all of those people and organizations that have helped make the production of this thesis possible. These include first of all my family and friends who have supported me, and my supervisor Dr. Rodolphe Devillers. My thesis co-advisor, Dr. Yvan Bédard (Université Laval) and my committee member, Dr. Olaf Schroth (UBC), are all appreciatively acknowledged. Additionally, I thank my colleagues in the Marine Geomatics Lab, Randal Greene, Andrew Cuff, Mel Agapito and René Enguehard, who helped with the prototypes, and Dr. Orland Hoeber for advice on the survey structure.

Also I would like to acknowledge the GEOIDE Network, Memorial University of Newfoundland, the Natural Sciences and Engineering Research Council of Canada (NSERC) and the NSERC Industrial Research Chair in Geospatial Databases for Decision Support for the funding that has made this research possible.

Finally I would like to thank the 40 anonymous participants who have completed the survey for the testing of my prototype, and the thesis reviewers.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1: Introduction

1.1 Context

The last decade has witnessed a significant trend towards the democratization of geospatial data and the development of new software and applications which utilize geospatial data. The democratization of geospatial data has been possible due, among other things, to public access to an increasing number of free geospatial data and applications. These applications include popular mapping websites, such as Google Maps and Bing Maps, which allow any Web user to visualise geospatial data. An increasing use of GPS and their associated applications in people's daily lives (e.g. car navigation systems, geocaching) (Rouse *et al.*, 2007) have also contributed to the democratization of geospatial data. Recent *Virtual Globe* (VG) software packages (e.g. Google Earth, Nokia Ovi 3D, Bing Maps 3D and NASA World Wind) allow any Internet user to view and sometime create geospatial data. Although there is no one universal or formal definition, Goodchild (2008) describes VGs as systems which provide access to huge sources of satellite imagery, aerial imagery, along with digital geospatial data, using a simple interface. Also, VGs provide an environment where many types of geospatial and contextual data can be viewed together or utilized for purposes ranging from simple to scientific in nature.

VGs are software whose capabilities have grown substantially in the past decade due mostly to technological advancements. They are commonly used for simple tasks, such as the visual identification of places, but can be used to support more complex decision making processes (Goodchild, 2008; Grossner *et al.*, 2008;

Sheppard and Cizek, 2009). With recent advancements in technology, more and more people are becoming users of VGs, and more and more data are available on the *geospatial web* (Haklay *et al.*, 2008). This influx of non-expert users creating and using geospatial data relates directly to the problem discussed by Dokoupil (2008) which asserts that, although experts seem to be pushed aside when it comes to creating web data, there is a realised danger in unreliable data on the web that people now see the need to address. Sheppard and Cizek (2009) address such outcomes in situations where experts fail to properly label scientific data, provide uncertainty information, and including relative contextual data, in visualizations used in VGs. Sheppard and Cizek (2009) also address dangers of error propagation and liability issues when images and visualizations created by non-expert users find their way into public discourse. For the purpose of this research, we define a non-expert user as someone with little or no training in or knowledge of the nature of geospatial data and an expert as someone who possesses formal training and practice using geospatial data, and these concepts can be seen discussed in greater detail in section 2.2.2.

In many applications and uses of geospatial data, the quality of the data should be taken into consideration. Zandbergen (2009) asserts that certain quality expectations have to be met when creating geospatial data. Sheppard and Cizek (2009) exemplify the importance of being aware of the authenticity of geospatial data such as 3D models, and the risks that may arise when these models find their way into being used in publicized images or incorporated into larger model sets. It is therefore important to know some measure of the quality of any geospatial dataset that one may use for decision making.

There are numerous measures of the quality of geospatial data. Beard and Mackaness (1993) state that spatial data quality has many definitions, extending from statistical to conceptual. The quality of geospatial data can be measured and described in many ways, although different professional standards (e.g. ISO 19113, Open Geospatial Consortium) have been developed to standardize these practices. The Spatial Data Transfer Standard (SDTS) outlines five main elements of spatial data quality: lineage, positional accuracy, attribute accuracy, logical consistency, and completeness (FGDC, 1991; Kresse & Fadaie, 2004). These criteria are sometimes referred to as *internal quality*. Other more recent spatial data quality standards (e.g. ISO 19113) followed similar criteria. Another view of geospatial data quality is the concept of *external quality*, which is also referred to as the data's *fitness for use*. This concept differs from the one of internal quality as it looks at "the level of fitness between data characteristics and user's needs" (Devillers *et al.*, 2005).

Internal and external data quality is typically communicated using statistical reports, disclaimers, user manuals, context-sensitive warnings, metadata, etc (Devillers & Beard, 2006). Another way to communicate data quality or uncertainty is through visualisation. MacEachren *et al.* (2005) argue that developing reliable methods to represent and manage data uncertainty is a persistent and relevant challenge in GISciences, and consequently efforts to develop tools and methods of visualisation for information uncertainty can be of benefit in helping analysts understand and cope with this uncertainty.

In this context, professionals and researchers from the GIS sector are considered as *experts*. For *non-expert users* of geospatial data, visualization may be a very useful way of communicating uncertainty and geospatial data quality (Buttenfield, 1993; MacEachren, 2005; Drecki, 2007). One potentially valuable area in which to implement these methods would be in 3D and VGs, as these new environments are growing in popularity and reaching increasing numbers of people that do not always understand the quality of the data which they are looking at.

The need to communicate geospatial data quality is now recognized as being important to the analysis of geospatial data (Roth, 2009) and to protect the geospatial data consumers (Gervais *et al.*, 2007). This communication is still more crucial as the access to geospatial data and the ability to create geospatial data becomes increasingly available to non-expert or *lay* GIS users via applications such as online mapping services (e.g., Google Maps, Bing maps, MapQuest) or VG environments (e.g., Google Earth, NASA World Wind, Bing Maps 3D, ESRI ArcExplorer).

1.2 Research problem

Lowell (2004) states that "decision making in human life involves a constant evaluation of risk". In order to evaluate these potential risks when using geospatial data for any type of decision making, one must be aware of the quality of the data. For instance, different 3D models obtained from Web sites like the Google 3D Warehouse can provide different representations of a same reality. A 3D environment visualized in a VG (e.g. a city) will potentially include models produced very differently by a number of users. The resulting quality may be very heterogeneous or even unknown. Having users be aware of this issue is one of the major challenges as much of this quality of geospatial data being used by non-expert GIS users is inherently unknown to users. This quality of the data may, depending on the data's use and/or analysis purpose, cause undesirable outcomes or unreliable results as Sheppard and Cizek (2009) exemplifies for geospatial data in landscape visualizations. With the advancements in technology, such as VGs and on-line mapping sites, geospatial data has become democratized. This leads to many new concerns with the need to communicate the quality of geospatial data being used and created by non-expert users so that they can understand what may result from the use of geospatial data for decision making.

Many of these non-expert users creating and using geospatial data are doing so using VGs. Due most likely to the recent apparition of VGs, few studies discuss users and uses of these systems (see section 2.2.2). A more comprehensive understanding of these user types is needed. Knowing who is using VGs is required to assess how to best communicate geospatial data quality to these end users.

Many ways in which geospatial data quality and uncertainty can be visualized have been explored (see section 2.3). This includes work from Drecki, 2002, 2007; MacEachren, 1992, 2005; Pang, 2001; Slocum *et al.*, 2005; and Zuk, 2008, which proposed visualizing data quality and uncertainty using methods such as likelihood surfaces, transparency levels, glyphs and colour. As of yet, there is no universally accepted 'best' or standard way to visually communicate geospatial data quality. Furthermore, of the visualization methods which have been previously proposed, none have been put into widespread use within the mainstream GIS technologies. With so many new non-expert GIS users, new challenges in determining geospatial data quality visualization methods are also presented. There is a need to assure that the visualization method is understandable for all users, not only expert users. One additional challenge related to the geospatial data quality visualization is also that the visualization method must be technically feasible for application to numerous and heterogeneous datasets. Balancing these practical constraints with the search for a theoretically sound solution has led to the adoption of the approach proposed in this thesis.

In summary, the research problem is that the quality of the geospatial data in VGs is currently not being communicated to users and that research is required to design approaches that will balance technical implementation constraints with powerful visualization that would allow raising the awareness about the quality of the data to users.

1.3 Research Hypothesis

The hypothesis to be tested by this research is that it is possible to visually communicate the geospatial data quality of 3D objects in VGs to non-expert users through a method which will be both technically feasible and easily understandable by non-expert users.

This hypothesis meets several qualities that a scientific hypothesis should have. It is *plausible* as methods for communicating quality to non-experts exist and are widely accepted in the non-spatial community of web users. It is *verifiable* by means of successfully building a working prototype and collecting feedback from a small group of representative users. It is *precise*, as the failure to develop the prototype during the short amount of time or the failure of users to understand the quality communicated would result in rejecting the proposed solution. Finally, the hypothesis is *communicable* as its statement can be easily understood.

1.4 Research Questions

The main research question for this research is "Is there an efficient and technically feasible method which can be used to communicate visually geospatial data quality associated to 3D objects in VGs to non-expert users of geospatial data?"

Specific research questions are:

- What are the different types of users of VGs?
- What methods for geospatial data quality visualization have been previously explored?
- Which method or set of methods can be used to communicate geospatial data quality of 3D objects to non-expert users of VGs?
- Which factors are important to consider in the development of an approach for implementing geospatial data quality visualizations for 3D objects in a VG?
- Which of the visualization method(s) implemented in the visualization prototypes is most effective for communicating the geospatial data quality to users?

1.5 Objectives

The goal of the research is to design, test and validate a *symbiotic approach* to communicate visually the perceived quality of 3D objects in VGs.

The concept of a *symbiotic approach*, which will be described in more depth in section 3.5.2, is an approach which balances the simplicity, the ease of implementation, and the detail of the visualization method in a trade-off which mutually benefits the users, developers and theoreticians of the method.

The specific objectives of this project are:

1) To identify and classify the different types of VG users
2) To identify and classify methods by which spatial data quality and quality more generally can be visualized
3) To design an approach for visualizing the quality of 3D geospatial data within VGs
4) To implement and test the approach in a visual prototype

1.6 Research Method

The method used for this project consists of five main phases (see Figure 1.1).

The first phase involved performing a literature review of the fields of geospatial data quality and uncertainty, uncertainty and quality visualization, Web 2.0, Volunteered Geographic Information, 3D geospatial environments, VG applications, and VG uses and users. This review helped in determining the current state of the field and identifying the research questions and the method to be used. The literature review also established the context into which the theoretical framework was set.

The second phase was to create an inventory of methods that can be used to visualize geospatial data quality and uncertainty. This was done through a study of the literature related to data quality and uncertainty visualisation, but also through a study of the ways in which quality of products is assessed and communicated in popular websites. This inventory provided a comprehensive overview of the current methods that can serve as a foundation for the theoretical framework.

The third phase was the development of a visual approach to communicate geospatial data quality to users of VGs. The approach taken looks at how an overall assessment of *Perceived Quality*, which will be defined in section 3.4.1, of 3D data can be visually conveyed in a VG environment. This approach also incorporates user feedback about certain quality aspects. Also, additional ways in which specific details of the uncertainty assessment can be communicated were examined (e.g. bar charts, user comments, quality slider). These techniques were used in a VG environment to communicate the geospatial data quality to potential users. The aim of this research was to provide specific information pertaining to individual 3D models (e.g. for each building) rather than information about the quality on a global scale for the whole dataset.

Phase four implemented the approach developed in phase three using different visual prototypes developed in Google Earth. User evaluations in the form of a survey of non-expert and expert users of geospatial data allowed gathering qualitative and quantitative data that can be used to validate or invalidate the research hypothesis. Many methods of data collection exist, such as direct observation, interviews and surveys. As the research requires understanding if users benefit from the communication method developed, it was decided that a formal user evaluation including a survey was the most appropriate form of data collection. To allow statistical tests to be significant, a group of 40 participants was surveyed, allowing the analysis and comparison of two sub-groups of 20 participants each (experts vs. non-experts). Participants were given a sample of 3D data in a VG with the quality visualised using different methods, and were asked to rank each of the visualization methods tested based on their ease of understanding, in addition to provide feedback based on their assessment. The outcome was quantitative and qualitative data which reflect the survey participants' understanding and assessment of the individual visualization prototypes.

Phase five of the research method was the analysis and discussion of the survey results. This allowed identifying the visualization method that was preferred by the users. The quantitative and qualitative results collected from the surveys were analysed. Participant's comments and evaluations provided qualitative insight for the approach validation. Participant's rankings allowed a quantitative analysis. Rankings of the visualization method were analysed using the Friedman's Test while differences in rankings between expert and non-expert participants were analysed using the Mann-Whitney Test.

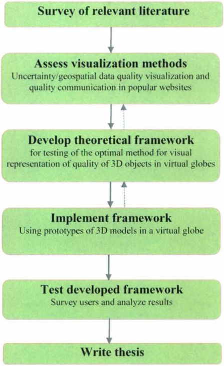

Figure 1.1: Research method summary

1.7 Conclusions

Due to the recent emergence and increasing popularity of VGs, online mapping and the ability for any user to create, access, and modify geospatial data, there are new challenges in communicating geospatial data quality to non-expert users. Most of the users of VGs and other on-line mapping applications typically have no expertise in geospatial data. Due to the heterogeneous nature of the data created and stored online, there are risks that users do not understand the quality of the data used to support different types of decisions. Hence, a communication method allowing non-expert users to understand easily the quality of these geospatial data is likely to reduce the risks of misusing these data.

The goal of this project is to design a method which is efficient, easy to understand and to implement, that could be used to communicate spatial data quality to non-expert VGs users. This is achieved through the development of a prototype which adapts visualization methods from popular websites to the VG's 3D environments. The prototype is then assessed through a user evaluation of both expert and non-expert users of geospatial data. The overall approach which is taken is a symbiotic one, which considers altogether pragmatic elements such as technical feasibility and ease of understanding along the theoretically best visual solution.

For this project a visual method was chosen for communicating data quality. Due to the lack of widespread use of previously explored methods for visually communicating geospatial data, methods from other fields which are non-spatial in nature were explored for use in this project.

1.8 Thesis organisation

This thesis follows a manuscript format wherein chapter 3 is a paper to be submitted to a peer reviewed journal. Chapter 2 is a literature review on VG users and usages, data quality and uncertainty visualization, Web 2.0 and geospatial data. Chapter 3 presents the new approach for visualizing perceived spatial data quality of 3D objects in VGs, which addresses the challenges and objectives identified in chapter 1. It presents an article that will be submitted to the international peer-review *Journal of Spatial Information Science*. This chapter presents the visualization approach, the prototype, and the results of a user survey on geospatial data quality visualization for determination of which of a subset of visualization methods is most effective. Chapter 4 presents and discusses additional results related to the visualization approach, key concepts, and the prototypes. Chapter 5 discusses the results and concludes the thesis. It explains how the hypothesis has been validated, and how the research questions have been answered. It also expands on how the research may support future research, and discussed the limitations of the work presented. A copy of the user survey is finally presented in Appendix A.

1.9 Co-authorship statement

This research is part of the larger Canadian GEOIDE IV 23 project titled "Public Protection and Ethical Geospatial Data Dissemination". The general scope of the research presented in this thesis (i.e. geospatial data quality in relation to VGs) was provided by my committee members. The specific research problem addressed, the objectives and the methods were developed by myself during the literature review. These ideas were then formalized in my thesis proposal with iterative reviews and feedback from my supervisor and co-supervisors. The practical aspects of this research, such as the development of prototypes and the execution and organisation of the user survey, were performed primarily by me. René Enguehard (MSc student) provided significant help with some of the technical aspects in the creation of KML files for the prototype. I have also completed all of the data analysis, which included the determination of the sub-set of visualization methods chosen for prototyping, and the analysis of the qualitative and quantitative data obtained from the user survey. I wrote the complete first draft, and am the primary author of the journal article presented in chapter 3, on which Rodolphe Devillers, Yvan Bédard and Olaf Schroth are co-authors. I am the author of this manuscript.

1.10 References

Beard, K., Mackaness, W., 1993. Visual Access to Data Quality in Geographic Information Systems. *Cartographica*, 30:2&3, pp. 37-47

Buttenfield, B. P., 1993. Representing Data Quality. *Cartographica*, 30:2&3, pp. 1-7

Devillers, R., Bédard, Y., Jeansoulin, R., 2005. Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS. *Photogrammetric Engineering & Remote Sensing*. 71 :2, pp. 205–215

Devillers, R., Beard, K., 2006. Communication and Use of Spatial Data Quality Information in GIS. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 237-254

Dokoupil, T., 2008. Revenge of the Experts. *Newsweek web exclusive*, March 6. Last accessed June 2011. <http://www.newsweek.com/id/119091>

Drecki, I., 2002. Visualization of Uncertainty in Geographical Data. In: Shi, W., Fisher, P. F., Goodchild, M. F., (Eds.), *Spatial Data Quality*. Taylor & Francis, New York: NY, pp. 140-160

Drecki, I., 2007. Geographical Information Uncertainty: The Concept and Representational Challenges. *Proceedings from the 23rd International Cartographic Conference*. Moscow, Russia, 13 p

FGDC (Federal Geographic Data Committee), 1991. *Spatial Data Transfer Standard*. Washington, DC.: Department of the Interior

Gervais, M., Bédard, Y., Jeansoulin, R., Cervelle, B., 2007. Qualité des données géographiques. Obligations juridiques potentielles et modèle du producteur raisonnable. *Revue Internationale de Géomatique*, 17:1, pp. 33-62

Goodchild, M. F., 2008. Spatial Accuracy 2.0. In: J.-X. Zhang and M. F. Goodchild (Eds.). Spatial Uncertainty. *Proceeding of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*. Volume 1. Liverpool: World Academic Union, pp. 1–7

Grossner, K. E., Goodchild, M. F., Clarke, K. C., 2008. Defining a Digital Earth System. *Transactions in GIS*, 12:1, pp. 145-160

Haklay, M., Singleton, A., Parker, C., 2008. Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*. 2:6, pp. 2011-2039

Kresse, W., Fadaie, K., 2004. *ISO Standards for Geographic Information*. Springer, Germany

Lowell, K., 2004. Why aren't we making better use of uncertainty information in decision-making? *Proceedings of the Joint Meeting of Accuracy 2004 and TIES 2004*. 7 p

MacEachren, A., 1992. Visualizing uncertain information, *Cartographic Perspectives*, 13, pp. 10-19

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, S., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32:3, pp. 139-160

Pang, A., 2001. Visualizing Uncertainty in Geo-spatial Data. In: *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*. pp. 1-14

Roth, R. E., 2009. The Impact of User Expertise on Geographic Risk Assessment Under Uncertain Conditions. Cartography and Geographic Information Science. 36:1, pp. 29-43

Rouse, L. J., Bergeron, S. J., Harris, T. M., 2007. Participating in the Geospatial Web: Collaborative Mapping, Social Networks and Participatory GIS. In: Scharl, A., Tochtermann, K., (Eds.). The Geospatial Web: How Geobrowsers, Social software and Web 2.0 are Shaping the Network Society. Springer-Verlag. London, pp. 153-159

Sheppard, S. R. G., Cizek, P., 2009. The ethics of Google Earth: Crossing thresholds from spatial data to landscape visualization. *Journal of Environmental Management*. 90, pp. 2102-2117

Slocum, T., McMaster, R., Kessler, F., Howard, H., 2005. *Thematic Cartography and Geographic Visualization*. Pearson Prentice Hall, NJ, 518 p

Zandbergen, P. A., 2009. Geocoding Quality and Implications for Spatial Analysis. *Geography Compass*. 3, p 1-34

Zuk, T. D., 2008. Visualizing Uncertainty. *PhD thesis*, Department of Computer Science, Calgary; Canada, 336 p

Chapter 2: Literature Review

2.1 Introduction

A review of the literature of several areas has been performed to assess the current state of several fields related to this project. The types of users and usages of VGs were first examined to allow determining what VGs are used for and the types of users. Following this, a review of the existing methods for visualizing geospatial data quality and uncertainty was done to understand the range of methods and to categorize them. The fields of Web 2.0 and VGI, which are important domains in relation to VGs and democratization of geospatial data, were also reviewed.

2.2 Virtual Globe Users and Usages

2.2.1 Overview of Virtual Globes

VGs have resulted largely from technological developments in areas such as tiling, caching and data transfer (Sheppard & Cizek, 2009) that occurred within the past decade. These developments along with the success of VGs have brought maps into the daily lives of people (Taylor & Caquard, 2006). VGs allow visualizing various types of geospatial data such as satellite and aerial imagery and vector data layers (Tuttle *et al.*, 2008), and some data are now able to be viewed in almost real-time (e.g. traffic data).

VGs have many applications and uses. According to Goodchild (2008) VGs provide access to huge sources of satellite imagery, aerial imagery, along with digital geospatial data, using a simple interface. He also conveys that this is done in an environment which can be rotated and zoomed to visualize the geospatial data without the distortions typically introduced through 2D maps. VGs integrate many types of geospatial and contextual data together to be viewed or utilized for purposes ranging from non-expert users simply viewing satellite data, to the creation of new data (e.g. creating a 3D building in Google Sketchup), or to more advanced purposes such as using VGs for scientific analysis.

The concept of a VG is attributed to the former US Vice-President Al Gore in 1998 who defined a visionary information system, or *digital earth*, with potential value for educational and collaborative research (Butler, 2006; Grossner *et al.*, 2008; Tuttle *et al.*, 2008). Although existing VGs do not include all of the requirements of Gore, they are comparable to his vision in many ways. This vision was partially realized with the initial launch of VGs such as NASA's World Wind in 2004 and Google Earth in 2005. It was different recent technological advancements, such as tiling techniques, which greatly reduce the size of file transfers (Editorial, 2006), that have allowed VGs to become a reality.

Today there are several VG software available. Most of these are available as freeware, open source, or with free licensing, although some VG providers offer upgraded versions of software at a cost. For instance, a professional version of Google Earth, which is aimed at organizations using VGs rather than the general public, is sold by Google. All VGs are not designed for the same purpose. Google Earth is designed for the use of the general public, whereas NASA's World Wind is "explicitly designed for scientific information and its code is open source so that scientists and software developers can tailor it to their needs" (Butler, 2006).

2.2.2 Virtual Globe Users

Due to the only recent emergence and popularity of VGs, there is still only a limited understanding of the breadth of their potential usage and users. Grossner *et al.* (2008) provide a first classification of VG users. They identify three primary categories of users from U.S. vice president Al Gore's speech: (1) non-expert users, which they exemplify as a young child visiting a digital museum, (2) collaborative scientists and (3) current GIS users, such as governments. Some examples of uses for each category can be seen in Table 2.1. While Grossner is to the best of our knowledge the only existing classification of VG users, a number of other studies proposed classification of users for the GIS environment, more specifically for collaborative environments. For example, Coleman *et al.* (2009) identify five categories of expertise, which overlap, pertaining to volunteered Geographic Information (VGI) contributors, a field related to VGs. These categories range from a *Neophyte*, who is a person with no background in the subject but has an interest or opinion, to an *Expert Authority*, who is a person which has greatly studied and practiced the subject and holds great authority in the area. For the purpose of this research, we define an expert in geospatial data as someone who possesses formal training and practice using geospatial data, and a non-expert as someone with no training or knowledge of the nature of geospatial data.

Table 2.1: Examples of VG uses from literature using Grossner *et al*., 2008 classification of users

Type of user	Example of use
Non-expert users	Navigate the streets of New York at eye level using Google Earth (Jones, 2007)
Collaborative scientists	Dissemination of object-based change detection research results using Google Earth (Tiede & Lang, 2007)
Current GIS users	Development of educational exercises using several virtual globes (Rakshit & Ogneva-Himmelberger, 2008)

In response to the poor knowledge of user types and, as part of the research for this thesis, a new classification of VG users is presented on Figure 2.1, and examples describing each of the four classes can be seen in Table 2.2. Creating this classification helped in achieving the objective of identifying and classifying the different types of VG users for this thesis.

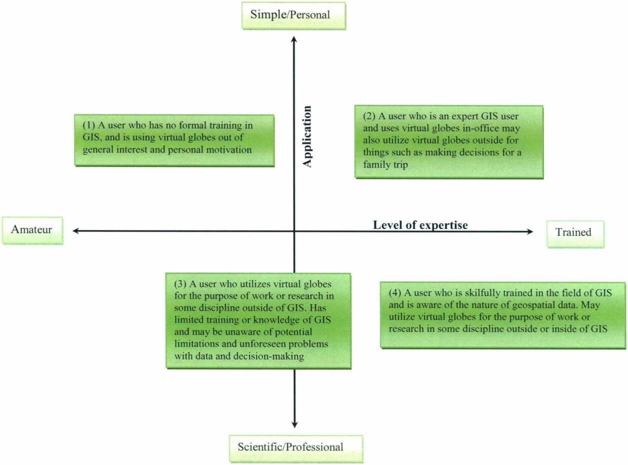

Figure 2.1: A new classification of virtual globe users

Table 1.2: The four classes of virtual globe users with examples

Classification	Example
(1) Amateur expertise – Simple/Personal application	Viewing satellite imagery of the road you live on
(2) Trained expertise – Simple/Personal application	Locate an address and navigate the streets around the address at eye level using Google Earth street view. (Jones, 2007)
(3) Amateur to Semi-trained expertise – Scientific/Professional application	The use of landscape visualizations using Google Earth by a Colorado group to visualize the skyline of the city (Sheppard & Cizek, 2009)
(4) Trained expertise – Scientific/Professional application	Dissemination of object-based change detection research results by researchers using Google Earth (Tiede & Lang, 2007). Analysis of population affected through fake crisis scenario in Google Earth [for European network on Global Monitoring for Stability and Security] (Tiede & Lang, 2007)

The proposed classification extends previous classifications by combining the complexity of the application with the level of expertise of the user. In a more general sense, VG users can be considered as experts, or non-experts in the use of geospatial data (see X axis in Figure 2.1). A person with little or no training in GIS or in the nature of geospatial data may be unaware of the potential limitations and uncertainties associated with the use of geospatial data for decision making. An expert may be a person with a formal training in GIS who understands the nature of geospatial data. Sheppard and Cizek (2009) also define experts and non-experts in the context of using VGs. Their definition is a more general one which does not incorporate expertise in GIS or geospatial data. They define experts as scientists and experts from various fields, being more from private sectors, and they define non-experts as the lay-public who are "people without particular expertise in science or environmental professions" (Sheppard & Cizek, 2009).

2.2.3 Uses of Virtual Globes

VGs are used in many different situations and for various reasons. It seems that many of these uses are everyday tasks. Taylor and Caquard (2006) convey that the success of these VGs in recent years has caused an increasing occurrence of map use in people's daily lives. Images from Google Earth are frequently now seen in news casts and TV shows to show and identify locations in the world. Other common uses of VGs include tasks such as road directions, locating buildings or sites or viewing potential travel destinations. Jones (2007), when speaking about Google Earth, highlights that people may also just explore places from history such as cultural, family or religious historical places.

Even though most everyday uses of VGs do generally not support analytical decisions, undesirable outcomes and ramifications may arise. For example, using a VG for routing may result in the identification of a bad route or wrong directions due to an incomplete dataset.

The following sections look at uses of VGs in analytical and educational uses, where these uses are relevant for collaborative scientists or for agencies that may be using VGs for decision making. It should be noted that some other potential commercial uses are not discussed in this review.

Analytical decision making uses

Analytical decision making, the dissemination of research to the public and responding to natural disasters speedily are possible uses of VGs (Editorial, 2006). Tiede and Lang (2007) outline some of the advantages of using VGs in conjunction with analytical results to disseminate data from Earth observation as "(1) the amount of data to be integrated is reduced; (2) this saves times in situations where rapid information delivery is required; (3) the rights of the original data are respected".

Examples where VGs are used in conjunction with, or for analytical decision making, include work from Tiede and Lang (2007), which describes one study where object-based change detection of urban areas in Harare was performed on Quick Bird imagery. Subsequently the analytical 3D view was implemented as a KML file in Google Earth. Another example which Tiede and Lang (2007) describe is one where 3D symbols representing the distribution of dwellings in refugee camps in an area of Tanzania were created using Google Sketchup. In this study, the data were then made accessible via a commercial GIS so that further GIS analysis in a VG could be allowed.

Other examples of VGs used in analytical decision making are outlined by Mangroves of Mexico (2010), Schroth *et al.* (2009), and Sheppard and Cizek (2009), in the context of landscape visualization. Sheppard and Cizek (2009) give the example of a visualization created in a VG which depicts the impact on the current landscape of forecasted sea level rise for part of British Columbia. This visualization was highly publicized but did not include any supporting scientific or uncertainty information that would allow showing a range of possible scenarios.

Educational uses

Another major use of VGs is for the purpose of education. Rakshit and Ogneva-Himmelberger (2008) describe several exercises in which VGs are used with pre-existing data for education in visualizing and analyzing spatial layers without the use of commercial GIS. The activities in the exercises they describe include:

estimating the extent of urban sprawl, visualizing internet use through time, toxic release inventory mapping, and an exercise on image classification and VGs. In these exercises three separate VGs were used.

Another example of education using VGs is illustrated in a course outline from the University of North Carolina. The course is used to educate on the "capabilities and limitations of Google Earth for regional archaeological survey and site location, and for general feature identification" (Madry & Henley, 2007). The course reflects previous work done in an area of France, where a large number of new and existing archaeological sites were recorded using Google Earth.

2.2.4 Conclusions

There are many types of VG users today, and the users' skill levels vary from high expertise with geospatial data, to none. The reasons for which people use VGs also greatly vary. For some, VGs are used in everyday tasks, such as planning a driving route, while in other cases scientists are using VGs to display and disseminate their work for analytical decision making. Further, VGs are also being used/proposed for educational purposes. In some cases, this is to analyze spatial layers without the use of costly software, and in other cases to train for decision making using VGs.

2.3 Data Quality and Uncertainty Visualization Techniques for 2D and 3D Geospatial Data

2.3.1 Introduction

Communicating data quality and uncertainty is important and problems related to quality and uncertainty in geospatial data are likely to increase with the growing interest in the geospatial web (Pang, 2001; Schlaisich *et al.*, 2004; Goodchild, 2008). Mainstream geospatial applications such as Google Earth provide access to large databases of geospatial data of heterogeneous quality but do not provide tools for communicating or visualizing data quality or uncertainty to the users (e.g. the date mentioned for the images visualized on Google Earth is not the acquisition date but the copyright date). MacEachren (1992, p. 10) states that "uncertainty is a critical issue in geographic visualization due to the tendency of most people to treat both maps and computers as somehow less fallible than the humans who make decisions they are based upon". There is then a need to communicate these issues to the users of the data.

Various methods exist for communicating uncertainty and quality in geospatial data, such as providing statistical output, text, metadata or geographic visualization. This section will focus on visualization approaches as a mean to communicate geospatial data quality and uncertainty. MacEachren *et al.* (2005) assert that developing reliable methods to represent and manage data uncertainty is a persistent and relevant challenge in GISciences, and consequently efforts to develop tools and methods of visualization for data uncertainty can be of benefit in helping analysts understand and cope with this uncertainty.

One reason why geospatial data quality and uncertainty visualization is important as a method for communicating uncertainty is that visualization may be the most effective communication method for non-expert users of geospatial data. Graphic depiction and drawing has historically been used to improve the study of statistical information (Buttenfield, 1993; Drecki, 2007). Also, there have been past demonstrations of non-expert users using visual representations of quality and uncertainty to better formulate decisions under indecision, rather than relying on statistical output. There are many statistical methods which professional analysts and researchers can use to attain measures of uncertainty, but many non-expert users are not capable of interpreting or using these methods (MacEachren, 2005; Roth, 2009). Bédard *et al.* (2007) relay that there are several cognitive sciences studies which show that using images instead of numbers or words is better for the stimulation of memory and understanding. Also visualization allows the user to see the variation in the uncertainty over space, which is not as easy to see through other means of communicating geospatial data quality and uncertainty (Drecki, 2002).

Although visualization is recognized as a valuable method for communicating quality and uncertainty in geospatial data, much is still needed in the ways of research in this field. There are still many challenges when it comes to communicating uncertainty in geospatial data; Zuk (2008, p. 33) conveys that: "The need for visualizing uncertainty along with data now has widespread acceptance. However the task of including the additional uncertainty information into an existing or new visualization while maintaining ease of comprehension for both the data and the uncertainty is not easy. As a result, the visualization of uncertainty is still not standard practice". Therefore, when designing a visualization method, it is crucial to consider how the ease of comprehension can be maintained, especially as many users of geospatial data are now non-experts.

2.3.2 Context

Spatial Data Quality and Uncertainty

Spatial data quality and uncertainty are terms often used as synonyms but that have different meanings. From a metadata perspective, spatial data quality refers to a number of elements described by various standards, such as the lineage, positional accuracy, attribute accuracy, logical consistency, and completeness (FGDC, 1991; Kresse and Fadaie, 2004). However, others associate the concept of spatial data quality to the idea of *fitness for use*, which looks at how datasets meet user's requirements (Devillers *et al.*, 2005).

Uncertainty on the other hand has also a number of formal definitions (Pang, 2001). The Merriam-Webster online dictionary defines uncertainty as: "(1) the quality or state of being uncertain: doubt (2) something that is uncertain" (Uncertainty, 2009). Uncertainty is a concept which is broadly known in many fields, geospatial uncertainty being a specific type. This may create confusion when trying to define uncertainty (Drecki, 2007). Drecki (2007) describes some of the various concepts of uncertainty in different fields of research as; in mathematics uncertainty is associated with possibility theory, in statistics uncertainty is associated with standard deviation, and in psychology uncertainty is associated with insecurity or expectation as part of the human condition. Zuk (2008) also points out that "Uncertainty is not isolated to statistical numerical processes but is a normal part of everyday life" in making decisions pertaining to things such as the weather or stock market. He also points out that uncertainty has benefits such as being used for privacy in photographs or video via the use of blurring effects. Overall uncertainty is a broadly encompassing concept which affects all fields of academia as well as everyday life.

Conceptual Models of Uncertainty for Geospatial Data

Several conceptual models of uncertainty related to geospatial data exist in the literature. To communicate the uncertainty present in geospatial data one must first understand the nature of uncertainty, and define the type of uncertainty to be visualized. This is important as the nature and type of uncertainty being communicated is essential to choosing the best technique for visualization.

Three of these conceptual models are Bédard's orders of uncertainty (1986), Fisher's model of uncertainty (1999), and Leyk's conceptual framework for uncertainty investigation in land-cover change modeling (2005). These three models were chosen to be discussed in this review as they give three very different perspectives on how uncertainty can be modeled depending on how one looks at the concept, and what the application or specific domain it is being used in. Bédard describes four orders of uncertainty which stem from conceptual to meta-uncertainty (see Table 2.3). Fisher, on the other hand, defines a hierarchy of uncertainty which stems from well defined and poorly defined objects (Figure 2.2). Leyk proposes a conceptual model consisting of three domains which encompass the main potential sources of uncertainty (see Table 2.4).

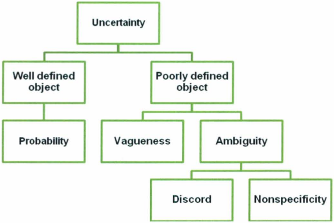

Figure 2.2: Conceptual model of uncertainty in geospatial data (Fisher, 1999)

Definitions of the sources of error in well defined and poorly defined objects are given by Fisher (1999). In a well defined object he states that uncertainty will be probabilistic and stem from errors. The uncertainty in a poorly defined object is labelled as vagueness if the object or objects class is poorly defined. Finally, the uncertainty in a poorly defined object is labelled as ambiguity if the object classification is ambiguous. If the object is shown as a member of two or more separate classes then the uncertainty is labelled as discord, and if the type of class the object should be in is not clear then the uncertainty is labelled as non-specificity.

Table 2.3: Levels of uncertainty and descriptions (Bédard, 1986 and 1988)

	Name	Description
First order	Conceptual uncertainty	Refers to the fuzziness in the identification of an observed reality (e.g. is it or is it not an entity? It is entity type A or B?)
Second order	Descriptive uncertainty	Refers to the vagueness in the attribute values of an observed reality (e.g. imprecision in quantitative values, unclear qualitative values)
Third order	Locational uncertainty	Refers to the vagueness in location in space and time of an observed reality (e.g. error ellipses in geodesy)
Fourth order	Meta-uncertainty	Refers to the degree to which the preceding uncertainties are known (e.g. absolute error ellipses with a probability of 39.3 percent)

Table 2.4: Definitions of uncertainty domains (Leyk, 2005)

Uncertainty Domain	Description
Production-oriented uncertainty	The amount of uncertainty inherent in the source data
Transformation-oriented uncertainty	The amount of uncertainty caused by data processing and editing
Application-oriented uncertainty	The amount of uncertainty dependent on the intended application

Theoretical context of visual variables

An important component of visualization is making proper use of visual variables. Basic visual variables were identified by Jacques Bertin as: Position (x,y), size, value, grain, color, orientation, and shape (Bertin, 1981; MacEachren *et al.*, 2005) (Figure 2.3). This list has since been built upon and modified (Slocum, 2005; MacEachren *et al.*, 2005). The additional visual variables added by MacEachren are contour crispness, fill clarity or resolution and transparency. The term *visual variables* is used to describe "the perceived differences in map symbols that are used to represent geographic phenomena" (Slocum, 2005). For the purpose of visualizing uncertainty, the visual variables identified by MacEachren *et al.* (2005), whom added additional variables pertaining to uncertainty, will be used. The additional variables identified in Slocum *et al.* (2005) (arrangement perspective height) will also be considered. For static 3D maps in particular, Häberling (2008) identifies five graphic variables; viewing inclination, zoom factor, light direction, haze density, and sky structure.

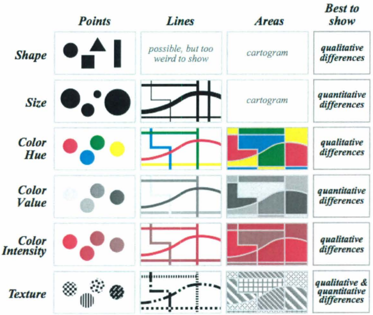

Figure 2.3: Bertin's visual variables (from: http://understandinggraphics.com)

2.3.3 Geospatial Data Quality Visualization Techniques

Visualization techniques can be broken down into various categories (e.g. static or dynamic, intrinsic or extrinsic), and techniques can be combined and used in different ways (e.g. color and size may be used together to represent different uncertainties).

Static and dynamic visualizations

Static or dynamic visualizations are two ways in which visualizations of geospatial uncertainty can be categorized. Davis and Keller (1997) assert that visualizations using static maps, or comparisons between static maps and visualizations using map animations, are two broad categories in which uncertainty visualization can be broken down. They also relay that, in general, for single static maps, the best variables to be used are color hue, color value and texture, and combinations of these. There are also situations in which map animations may be of benefit for uncertainty visualization. MacEachren *et al.* (1997) assert that map animations may be very important in fields such as health care for monitoring diseases over time.

Using a wireframe to represent parts of an object is one technique which exemplifies a static method of geospatial data quality and uncertainty visualization. Zuk (2008) gives an example of this method where an archaeological object being reconstructed is shown to be composed of wireframe in less certain areas. Another example of a static visualization method is given by Pang (2001) where breaks in contour lines show the uncertainty in the data (see Figure 2.4).

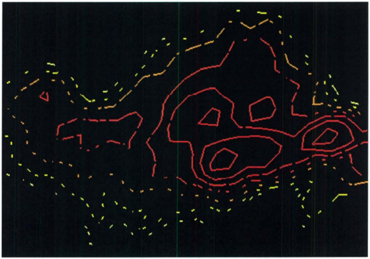

Figure 2.4: Use of gaps in contour lines to communicate uncertainty in the dataset (Pang, 2001)

The use of a moving/interactive timeline is one way in which geospatial uncertainty can be visualized dynamically. Zuk (2008) illustrates one such example with 3D archaeological data where moving the timeline creates haze on the image depending to indicate the level of uncertainty (see Figure 2.5). This example is extrinsic and on-screen.

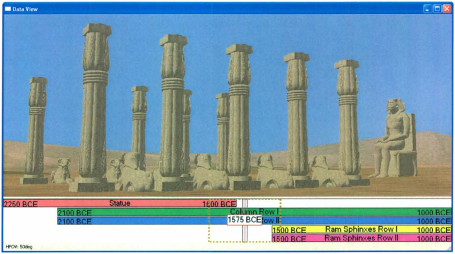

Figure 2.5: The use of a dynamic timeline to communicate geospatial data quality and uncertainty visualization (Zuk, 2008)

The use of intrinsic or extrinsic visualizations

One way in which visualizing geospatial uncertainty can be differentiated is by whether the technique uses intrinsic or extrinsic variables. Intrinsic geospatial data quality and uncertainty visualizations are inherent to the data display (MacEachren *et al.*, 2005; Slocum *et al.*, 2005). An intrinsic visualization of geospatial uncertainty would be to depict the uncertainty through transparency, where the data are used to display the uncertainty. Extrinsic geospatial data quality and uncertainty visualizations occur when visual variables are added to the map or environment and are not an integral part of the data (MacEachren *et al.*, 2005; Slocum *et al.*, 2005). For example, extrinsic geospatial data quality and uncertainty visualization may be when additional symbols are added to a map to represent uncertainty instead of modifying the existing data on the map. In a study conducted by MacEachren *et al.* (2005), it was found that the use of intrinsic techniques was better for communicating the "big picture" about geospatial uncertainty, whereas the use of extrinsic techniques was found to be more successful in geospatial uncertainty information from specific locations. An example of this would be a dashboard display of uncertainty or the use of glyphs.

The use of transparency on an object can be used for intrinsic geospatial data quality and uncertainty visualization. Zuk (2008) gives an example of this where the transparencies of the walls of a 3D house convey geospatial uncertainty (Figure 2.6).

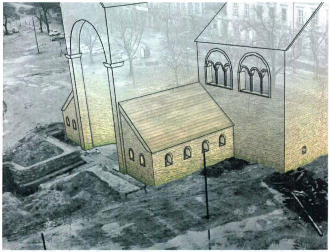

Figure 2.6: Use of transparency to convey geospatial uncertainty on a 3D building (Zuk, 2008)

The use of glyphs to communicate geospatial data quality and uncertainty is an example of extrinsic visualization. Pang (2001) gives the example of a vector field of glyphs showing uncertainty in the orientation of the glyph, and the amount of uncertainty in the size (Figure 2.7). This example is both on-screen and static. Riazanoff and Santer (2006) also provide an example where error vector glyphs are used to convey the localized spatial data quality of a raster map.

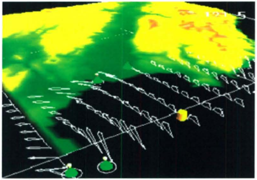

Figure 2.7: Vector field of glyphs showing geospatial uncertainty (Pang, 2001)

The use of visualizations such as quality indicators for logical consistency and thematic accuracy is another extrinsic visualization method. Devillers and Beard (2006) provide an example illustrating this method using a quality dashboard.

Separate screen vs. on-screen visualization

Geospatial data quality and uncertainty visualization can also be broken down into separate screen or on-screen. Separate screen uncertainty visualizations occur when the uncertainty visualization is not a part of the map or environment itself, but displayed on a separate screen. On-screen visualizations occur when the uncertainty visualization is a part of the map or environment where the dataset is displayed. MacEachren *et al.* (2005) describe a case study where two of these methods were used in a comparison. The on-screen geospatial data quality and uncertainty visualization had the pixels on the map more highly saturated in color for more certain areas. The separate screen visualization had an animated map flickering back and forth between the original data map and one which showed only pixels with high certainty. Ehlschlaeger *et al.* (1997) also speak about separate screen visualization of geospatial data. They assert that using animations to visualize geospatial uncertainty "can be an invaluable aid for exploratory analysis of the data".

The display of two side by side maps is one way in which separate screens are used to visualize geospatial data quality. Devillers and Beard (2006) exemplify this method through side by side raster maps, one showing the original data, and the other showing the associated uncertainty in the same area. MacEachren (1992) also gives an example of this where a raster map depicting a health risk index is shown next to a map of the same areas depicting the uncertainty in the health risk data (Figure 2.8). This example is also both static and extrinsic.

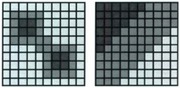

Figure 2.8: Use of separate screens for the visualization of geospatial uncertainty where a health risk index for air pollution is on the left and the uncertainty of those index values is shown on the right (MacEachren, 1992)

Making certain areas fuzzy or out of focus is a way in which on-screen geospatial uncertainty can be visualized. MacEachren (1992) shows an example of this as making the risk zone area surrounding a nuclear power plant more out of focus in less certain areas (Figure 2.9). This example is also both static and intrinsic.

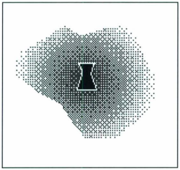

Figure 2.9: The use of out of focus raster data to communicate geospatial uncertainty where the more out of focus the data is in the zone, the less certain the data is (MacEachren, 1992)

Individual methods of geospatial data quality visualization

Many combinations of visual variables and applications in 2D and 3D were found by observing some of the individual methods for geospatial data quality visualization in the literature. To summarize some of these methods, a table using Bertin's (Bertin, 1981; MacEachren *et al.*, 2005; Slocum, 2005; Zuk, 2008), MacEachren *et al.*'s (2005), and Slocum *et al.*'s (2005) visual variables was created as part of this research (See Table 2.5). This table is categorized by variable, and indicates whether the variables are used for 2D or 3D, and provides examples of the use of each variable in geospatial data quality. The information on whether Bertin's visual variables (and some additional ones) are usable in 2D and 3D was found in Slocum *et al.* (2005). The usability of other visual variables in 2D or 3D is determined on the basis of known examples. Researching the literature and creating this table helped in achieving part of the second objective of this thesis by identifying and classifying methods by which spatial data quality can be visualized.

Table 2.5: Table illustrating examples of geospatial data quality uncertainty visualization by visual variable

Visualization variable	Used in 2D	Used in 3D	Example(s)	Data types variable is appropriate for	Dynamic visualization	Color blind ok
Location	Yes	Yes	Floating and sinking objects (Zuk, 2008),	Raster, 3D , points, lines, polygons	Yes	Yes
Size	Yes	Yes	Tissot indicatrix (Slocum *et al.*,2005), Line width depicting uncertainty (Slocum *et al.*, 2005), Use of glyphs (Pang, 2001)	3D, points, lines, polygons	Yes	Yes
Color value	Yes	Yes	Diverging color scheme depicting uncertainty (MacEachren *et al.*, 2005)	Raster, 3D, points, lines, polygons	Yes	No
Grain	Yes	Yes	Using interpolation to show certainty in boundaries (Davis & Keller, 1997)	Raster, 3D	Yes	Yes
Color hue	Yes	Yes	Varying hues to represent data age (Schlaisich *et al.*, 2004)	Raster, 3D, points, lines, polygons	Yes	Yes
Orientation	Yes	Yes	Use of glyphs (Pang, 2001)	Raster, 3D, lines, polygons	Yes	Yes
Shape	Yes	Yes	Tissot indicatrix (Slocum *et al.* 2005)	3D, points, lines, polygons	Yes	Yes
Color saturation	Yes	Yes	Color saturation of individual pixels to indicate accuracy (Schlaisich *et al.*, 2004)	Raster, 3D, points, lines, polygons	Yes	No
Contour crispness	Yes	Yes	Line fuzziness indicating uncertainty of boundary (Slocum *et al.*, 2005),	Raster, 3D, points, lines, polygons	Yes	Yes
Fill clarity\ Resolution	Yes	Yes	Fuzzy classification of imagery data (MacEachren *et al.*, 2005), Rain/snow and fog/haze techniques (Zuk, 2008)	Raster, 3D	Yes	Yes
Transparency	Yes	Yes	Transparency of uncertain 3D artifacts (Zuk, 2008),	Raster, 3D, points, lines, polygons	Yes	Yes
Spacing (texture)	Yes	Yes	Spacing between iso lines to show weather model uncertainty (MacEachren *et al.*, 2005)	Points, lines, polygons	No	Yes
Perspective height	Yes	No	Using a 3D graph to show the quality attributes of an area (e.g. resolution) (Schlaisich *et al.*, 2004)	Raster, points, lines, polygons	No	Yes

2.3.4 2D vs. 3D Geospatial Data Quality Visualization

Many of the visual variables, or combinations of variables used in geospatial data quality visualization techniques, can be applied to both 2D and 3D representations. As there were no known case studies which compared or contrasted specific visualization methods between 2D and 3D, it is difficult to compare specific techniques between the two. Visualizing geospatial data quality in 3D environments may pose more challenges as to how certain visual variables may be used. For example, size may be used, but not the size of multiple 3D objects as distance would change the viewers' perception of the objects size, rather the size of symbols overlaid on a single object might be used to depict uncertainty.

2.3.5 Comparison of Geospatial Data Quality Visualization Techniques

Overall, each of the techniques for visualizing geospatial data quality has been found useful. The way in which they seem to be distinguished is through their degree of usefulness, and this depends on the type of uncertainty which is trying to be communicated, the user, and the use. According to Roth (2009) "The key to designing useful and usable representations and visualizations of uncertainty (or of anything else) is to know the end user". Aside from the uncertainty type, the use is also important. As MacEachren *et al.* (2005) found, depending on whether the uncertainty to be communicated is part of the big picture, or in specific parts of the dataset, an intrinsic or extrinsic approach may be more useful. For this a good understanding of the nature and types of uncertainty could be helpful. This may be achieved through the use of conceptual models of uncertainty, such as those seen in section 2.2.

2.3.6 Conclusions

There are a broad range of methods and techniques through which geospatial data quality can be visualized. These include the many ways in which quality visualization may be broken down (e.g. intrinsic or extrinsic), and also the many visual variables and combinations of those which may be utilized.

One thing to consider is whether the visualization will be applied to 2D or 3D data. There are many more examples of uncertainty visualization for 2D datasets then there are for 3D. One recommendation would be to further investigate the usefulness of quality visualization techniques applied to 3D data. Jones (2007) states that "Some view Google Earth as the most widely used 3D visualization tool ever created". This, combined with the fact that most people using tools such as Google Earth with 3D data are non-experts in GIS which may not understand other ways of uncertainty communication (e.g. statistical output), exemplifies the importance of more research into techniques for uncertainty visualization using 3D data.

2.4 Web 2.0 and Geospatial Data

2.4.1 Introduction

Web 2.0 is an integral reason for the democratization of geospatial data and how any internet user can now create, use and access geospatial data. Davidson and Vaast (2007) state that Web 2.0 "harnesses the Web in a more interactive and collaborative manner, emphasizing peers' social interaction and collective intelligence, and presents new opportunities for leveraging the Web and engaging its users more effectively." Goodchild relays that Web 2.0 is a concept which relates to the users large involvement in creating and contributing content on the web (2008). He also conveys that the Web 2.0 concept has importance in geospatial information. Some examples of Web 2.0 applications are *wikis* (e.g. Wikipedia), social networking sites (e.g. Facebook), data sharing/peer to peer sites (e.g. alluc.org), VGs, and online mapping sites (e.g. Open Street Map).

2.4.2 Volunteered Geographic Information

VGI is another component of the Web 2.0. The growth and emergence of VGI is largely attributed to the growth of Web 2.0 applications in recent years (Nkhwanana, 2009). With this growth in VGI, questions related to the quality and uncertainty of the data produced by non-expert users of geospatial data are becoming more pressing. Goodchild (2008) conveys that there are many questions as to the accuracy of VGI data as some sites allow any user to edit the data.

The concept of VGI is especially seen in online mapping applications. Sites such as Open Street Maps are run on the basis of non-expert users adding and building upon the street maps which exist. Although the data on such online mapping sites is VGI, it is also subject to a peer reviewing process before it is used in the map. VGI is also encountered in VG applications. One example of this is Google Earth, where anyone can add information about a geographic area (e.g. pictures, descriptions, points). Google

Earth also provides the ability for users to create 3D models of buildings which are then freely available through their 3D Warehouse for any person to use or view.

Many other examples of VGI are seen through websites documenting various types of phenomena, from animal sightings to documenting where certain weather conditions occur. One example is the Web site www.junponline.com, where users provide information such as pictures, descriptions, and locations of plant and wildlife sightings to a larger inventory/database.

2.4.3 Conclusions

Much of the ways in which non-expert users understand and use geospatial data is related to Web 2.0. Applications and sites which facilitate peer to peer sharing and VGI, such as VGs and mapping sites, seem to be the main outlets by which non-experts are using and creating geospatial data. Although there seems to be no current way in which the quality of all VGI is being assessed, in many Web 2.0 related sites there are ways in which the quality of both products and providers of non-geospatial services are being assessed through peer-review processes such as user ratings and user feedback.

2.5 References

Bédard, Y., 1986. A Study of Data Using a Communication based Conceptual Framework of Land Information Systems. *Le Géomètre Canadien*. 40:4, pp. 449-460

Bédard, Y., 1988. Uncertainties in Land Information Systems Databases. *AUTO-CARTO* 8. pp. 175-184

Bédard, Y., Rivest, S. & Proulx, M. J., 2007. Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. Data Warehouses and OLAP: Concepts, Architectures and Solutions. In: Wrembel, R., Koncilia, C. (eds.), *Data Warehouses and OLAP: Concepts, Architectures and Solutions*. IGI Global. pp. 298-319

Bertin, J., 1981. *Graphics and Graphic Information-Processing*. Walter de Gruyter, New York

Butler, D., 2006. The web-wide world. *Nature*, 439:7078, pp. 776-778

Buttenfield, B. P., 1993. Representing Data Quality. *Cartographica*, 30:2&3, pp. 1-7

Coleman, D. J., Georgiadou, Y., Labonte, J., 2009. Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*. 4, pp. 332-358

Davidson, E. Vaast, E., 2009. Tech Talk: An Investigation of Blogging in Technology Innovation Discourse. Professional Communication, *IEEE Transactions*. 52:1, pp. 40 - 60

Davis, P., Keller, C., 1997. Modeling and Visualizing Multiple Spatial Uncertainties. *Computers & Geosciences*. 23:4, pp. 397-408

Devillers, R., Bédard, Y., Jeansoulin, R., 2005. Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS. *Photogrammetric Engineering & Remote Sensing*. 71:2, pp. 205–215

Devillers, R., Beard, K., 2006. Communication and Use of Spatial Data Quality Information in GIS. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 237-254

Drecki, I., 2002. Visualization of Uncertainty in Geographical Data. In: Shi, W., Fisher, P. F., Goodchild, M. F., (Eds.), *Spatial Data Quality*. Taylor & Francis, New York: NY, pp. 140-160

Drecki, I., 2007. Geographical Information Uncertainty: The Concept and Representational Challenges. *Proceedings from the 23rd International Cartographic Conference*. Moscow, Russia, 13 p

Editorial. 2006. Think global. *Nature*, 439:7078, pp. 763

Ehlschlaeger, C. A., Shortridge, A. M., Goodchild, M. F., 1997. Visualizing Spatial Data Uncertainty using Animation. *Computers & Geosciences*. 23:4, pp 387-395

FGDC (Federal Geographic Data Committee), 1991. *Spatial Data Transfer Standard.* Washington, DC.: Department of the Interior

Fisher, P., 1999. Models of Uncertainty in Spatial Data. In: Longley, P., Goodchild, M. F., Maguire, D. J., Rhind, D. W., (Eds.), *Geographical Information Systems*. John Wiley, New York: NY, pp. 191-205

Goodchild, M. F., 2008. Spatial Accuracy 2.0. In: J.-X. Zhang and M. F. Goodchild (Eds.). Spatial Uncertainty. *Proceeding of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*. Volume 1. Liverpool: World Academic Union, pp. 1–7

Grossner, K. E., Goodchild, M. F., Clarke, K. C. 2008. Defining a Digital Earth System. *Transactions in GIS*, 12:1, pp. 145-160

Häberling, C., 2008. Proposed Cartographic Design Principles for 3D Maps: A Contribution to an Extended Cartographic Theory. *Cartographica: The International Journal for Geographic Information and Geovisualization* 43, 175-188

Jones, M., 2007. Google's Geospatial Organizing Principle. *IEEE Computer Graphics and Applications*. 27: 4, pp. 8-13

Kresse, W., Fadaie, K., 2004. *ISO Standards for Geographic Information*. Springer, Germany

Leyk, S., Boesch, R., Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modeling. *Transactions in GIS*, 9:3, pp. 291-322

MacEachren, A., 1992. Visualising uncertain information. *Cartographic Perspectives*, 13, pp. 10-19

MacEachren, A., Polsky, C., Haug, D., Brown, D., Boscoe, F., Beedasy, J., Pickle, L., Marrara, M., 1997. Visualizing Spatial Relationships Among Health, Environmental, and Demographic Statistics: Interface Design Issues, *Proceedings of the 18th International Cartographic Conference*. Stockholm, Sweden, 9 p

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, S., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32:3, pp. 139-160

Madry, S., Henley, A., Google Earth for Scientific Regional Analysis. 2007. University of North Carolina. July 7, 2009. <http://www.lib.unc.edu/reference/gis/ge>

Mangroves of Mexico as Google Earth Outreach example for GE6 in cooperation with CONABIO. 2010. In: *Landscape Visualization*. Last accessed, June 2011 <http://www.lviz.org/?p=349>

Nkhwanana, N. J., 2009. Assessing the Credibility and Trust of Volunteers of VGI. *White paper prepared for GEOIDE project IV- 41*. 23 p

Pang, A., 2001. Visualizing Uncertainty in Geo-spatial Data. In: *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*. p. 1-14

Rakshit, R., Ogneva-Himmelberger, Y., 2008. Application of Virtual Globes in Education. *Geography Compass*. 2:6, pp. 1995-2010

Riazanoff, S., Santer, R., 2006. Quality of Raster Data In Spatial Data. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 43-60

Roth, R. E., 2009. The Impact of User Expertise on Geographic Risk Assessment Under Uncertain Conditions. *Cartography and Geographic Information Science*. 36:1, pp. 29-43

Schlaisich, I., Mountrakis, G., Agouris, P., 2004. Visualization of Image Quality in Distributed Spatial Databases. In: M. O. Altan (Editor), *International Society for Photogrammetry and Remote Sensing XXth Congress*. Istanbul, Turkey

Schroth, O, Pond, E., Muir-Owen, S., Campbell, C., Sheppard, S., 2009. Tools for the understanding of spatio-temporal climate scenarios in local planning: Kimberley (BC) case study. *SNSF Report PBEZP1-122976*. 33 p

Sheppard, S. R. G., Cizek, P., 2009. The ethics of Google Earth: Crossing thresholds from spatial data to landscape visualisation. *Journal of Environmental Management*. 90, pp. 2102-2117

Slocum, T., McMaster, R., Kessler, F., Howard, H., 2005. *Thematic Cartography and Geographic Visualization*. Pearson Prentice Hall, NJ, 518 p

Taylor, F., Caquard, S., 2006. Cybercartography: Maps and Mapping in the information Era. *Cartographica*. 41:1, pp 1-5

Tiede, D., Lang, S., 2007. Analytical 3D Views and Virtual Globes – Putting Analytical Results into Spatial Context. *Paper from: ISPRS Joint Workshop "Visualization and Exploration of Geospatial Data"*. 6 p.

Tuttle, B. T., Anderson, S., Huff, R., 2008. Virtual Globes: An Overview of Their History, Uses, and Future Challenges. *Geography Compass*. 2:5, pp. 1478-1505

Uncertainty. 2009. In: *Merriam-Webster Online Dictionary*. Last accessed, June 2011 <http://www.merriam-webster.com/dictionary/uncertainty>

Zuk, T. D., 2008. Visualizing Uncertainty. *PhD thesis*, Department of Computer Science, Calgary; Canada, 336 p

Chapter 3: Visualizing perceived spatial data quality of 3D objects within virtual globes

3.1 Introduction

The emergence of Virtual Globes (VGs) in the five past years has significantly changed the geographic information landscape by giving an easy access to any Web user to a very large volume of geographic data. In defining a VG, Rakshit and Ogneva-Himmelberger (2008) describe numerous aspects of the technology. They define a VG as a globe having the capability to simultaneously represent many thematic views of the Earth's surface. Additionally, they define a VG as an environment where the user can zoom, rotate, and tilt their view, and can display many types of data. VGs integrate satellite (raster), vector, and other types of contextual geospatial data together to be viewed or utilized for many purposes. These purposes range in use from non-expert users simply viewing Earth imagery, to the creation of new data to be viewed in, or integrated with the VGs data (e.g. creating a 3D building in Google SketchUp), to people utilizing VGs for scientific analysis (e.g. Butler, 2006; Tiede & Lang, 2007; Sheppard & Cizek, 2009; Hoeber *et al.*, 2010).

The larger vision for a digital Earth suggested in 1998 by the former US Vice-President Al Gore (Tuttle *et al.*, 2008) was partially realized when NASA's World Wind was launched in 2004, shortly followed in 2005 by Google Earth. Currently there are several VGs available for use on-line, including Google Earth, Nokia Ovi 3D, Bing Maps 3D, NASA World Wind, and Marble. This technology has enabled lay Internet users to integrate, display, analyse, create and share geospatial data. The sharing and creation of geospatial data on-line named Volunteered Geographic Information (VGI), is another new phenomenon which has contributed to the democratization of geospatial data and mapping.

The democratization of geospatial data was also supported by the emergence of what is defined as the Web 2.0, which promoted the development of web applications supporting VGI. While the production of geographic information used to be restricted to mapping professionals, anyone who has access to the Internet can now contribute to the creation of public geographic datasets through applications such as OpenStreetMap[1].

The Web 2.0 and VGs have allowed for many new opportunities in the geographic community, such as on-line collaboration, data sharing and access to many new tools and data. They have however presented new challenges to the research community. Coleman (2010) and Coleman *et al.* (2009) characterize VGI contributions of individuals as being either constructive or damaging and relays that one of the major challenges pertaining to VGI is assessing the credibility of these contributors. Another challenge related to VGI is the assessment of the quality of these data. Haklay (2010) examines this challenge through a comparative study of the OpenStreetMap and Ordnance Survey datasets in UK. He concludes that the quality is not unlike that of government datasets, but the distribution of error in the OpenStreetMap dataset is related to the contributor, and therefore not randomly distributed as expected in a governmental

---

[1] http://www.openstreetmap.org/

or commercial dataset. Girres and Touya (2010) extend this work through examining the OpenStreetMap data in France.

These challenges in assessing the quality and credibility of VGI are also seen in VGs where user-generated 3D models are being used. Many of these 3D models, such as those found in the Google Warehouse and in Google Earth, do not come with quality information or assessments by experts. Therefore, there is a need to design a method for assessing the geospatial data quality of these models and communicating this quality to the people using these models. Since a lot of the people using VGs are non-experts in the domain of geospatial data, communicating the quality of this data becomes even more challenging as many traditional methods may not be easily understandable by non-expert users.

As many users of VGs are non-experts in geospatial data, risks of misuse may exist when users utilize those 3D models to support decision making processes. Sheppard and Cizek (2009) discuss these risks in the context of landscape visualization. These risks include misinformation being published or unofficial images or data entering into decision making. One example of this is a flood map of part of British Columbia, Canada, created by the Sierra Club of British Columbia which became highly publicized although it was not accompanied by uncertainty information or substantial scientific framing (Sheppard & Cizek, 2009). Another set of potential risks Sheppard and Cizek highlight is the creation of inaccurate 3D models by non-experts which are then used in an official model and then may cause error propagation and potentially even liability and legal issues.

The goal of this research is to communicate the quality of 3D objects in VGs to users, which may help reduce the risks related to decision making based on these models. It aims to design, test, and validate a symbiotic approach to communicate visually the perceived quality of 3D objects in VGs. We first present an overview of spatial data quality (SDQ) and VG users and usages. Subsequently, an approach for the visualization of perceived SDQ of 3D objects within VGs is presented, along with the novel concept of perceived quality, the type of SDQ communicated by the approach. The prototyping of this approach and the testing of the method via a user evaluation are then discussed, followed by the conclusions.

3.2 Spatial Data Quality

3.2.1 Overview

Spatial Data Quality (SDQ) is a term which has many definitions extending from statistically measurable error and accuracy to conceptual measures of completeness and consistency (Beard and Mackaness, 1993). Five main elements of spatial data quality are outlined by the Spatial Data Transfer Standards (SDTS): lineage, positional accuracy, attribute accuracy, logical consistency, and completeness (Chrisman, 1983; FGDC, 1991; Kresse & Fadaie, 2004). Another view of SDQ lies in the concept of fitness for use, also named external quality. Devillers and Jeansoulin (2006) convey that external quality is generally acknowledged as quality in the broadest sense. Also, external quality is not fixed, in the sense that one dataset may have different qualities for different end users.

SDQ communication is an active area of research in the field of SDQ as there currently are no methods which are in widespread use. Communicating SDQ issues is currently mostly done through metadata, although a number of visualization methods have been proposed. Metadata is often defined as 'data about data'. Examples of metadata include the projection of a dataset, its spatial and temporal extent, or its spatial accuracy. Several problems come with the use of metadata to communicate SDQ issues (Devillers *et al.*, 2005). Since metadata are not mandatory, they are not always created along with the dataset, and therefore not available to users. Another problem is that metadata do not all follow a same standard. The International Organization for Standardization (ISO) is generally recognized as the standard, but still not universally. Metadata are also usually not consulted by many GIS users as they are normally disjointed from datasets and remain fairly complex.

Another concern with using metadata is that information conveyed in metadata often describes the entire dataset and not individual objects. Drecki (2002) points out that the overall picture of the dataset is often satisfactory, but that this does not mean that data quality is evenly distributed within the dataset. Hunter and Goodchild (1996) assert that traditional ways of conveying uncertainty have been global, referring to the quality of the hard copy map in general.

Statistical methods can also be used to assess and describe SDQ. Beard and Mackaness (1993) relay that the root mean squared error (RMSE) is an aspatial statistical measure which provides a summary measure of positional accuracy for a dataset. This measure, along with many other statistical measures, is of a global nature. This distinction becomes important when considering visualization as a technique for communicating data quality, as using a spatial visualization technique can display local versus global variation (Beard and Mackaness, 1993). Also, MacEachren *et al.* (2005) examine the validity of the visual representation of data uncertainty as there are many statistical methods which professional analysts and researchers can use to attain measures of uncertainty, but many lay users are not capable of interpreting or using these methods. Visualizing SDQ has the advantage of being able to display local variation in data quality. The research of MacEachren *et al.* (2005) also highlights that there have been past demonstrations of lay users depending on heuristics to formulate decisions based on uncertainty, rather than relying on statistical output. Parallel to this, Buttenfield (2003) exemplifies depicting error and confidence intervals graphically for exploratory data analysis in statistical analysis. She asserts that graphically depicting or drawing statistical data has improved the study this information, which could similarly improve the depiction of error and uncertainty.

3.2.2 Analysing Visualization Methods for Geospatial Data Quality

Many methods have been proposed to visualize geospatial data quality. Some examples are the use various colour schemes, opaqueness, or fuzziness to represent the SDQ of the data (Schlaisich *et al.*, 2004; MacEachren *et al.*, 2005; Zuk, 2008). Visualization methods for communicating geospatial data quality can be classified in various ways. Generally, these methods use traditional visual variables (e.g. colour, size, transparency) to portray the quality of the data. Much previous work focused on 2D mapping, although there also has been some work for 3D environments (e.g. Zuk, 2008).

Visualizations can be classified as either static or dynamic. Intrinsic or extrinsic variable can also be used, where intrinsic geospatial data quality and uncertainty visualizations are inherent to the data display (MacEachren *et al.*, 2005; Slocum *et al.*, 2005), and extrinsic variables are those added to the map or display. Rather than visualizing the geospatial data quality and uncertainty on the same map or screen, a separate map or screen may be used to display the visualization.

Very few methods for visualizing the geospatial data quality of 3D objects in VGs have been proposed. One recent study by Peter (2009) presents a method for evaluating inconsistencies in 3D building models. This method may be useful for visually examining and identifying inconsistencies, but it is one which relies on very detailed models for comparison, whereas the method developed in this research is one which utilizes users' perceptions and knowledge of the 3D objects to visualize quality.

Many of these visualization approaches can be used in combinations to visualize geospatial data quality. Table 3.1 classifies methods proposed in the literature according to 13 visual variables. This served as an essential step in understanding what potential methods or aspects of methods could be implemented in visualizing geospatial data quality for 3D objects within a VG environment.

Table 3.1: Table illustrating examples of uncertainty visualization by visual variable

Visualization variable	Used in 2D	Used in 3D	Example(s)
Location	Yes	Yes	Floating and sinking objects (Zuk, 2008),
Size	Yes	Yes	Tissot indicatrix (Slocum *et al.*,2005), Line width depicting uncertainty (Slocum *et al.*, 2005), Use of glyphs (Pang, 2001)
Color value	Yes	Yes	Diverging color scheme depicting uncertainty (MacEachren *et al.*, 2005)
Grain	Yes	Yes	Using interpolation to show certainty in boundaries (Davis & Keller, 1997)
Color hue	Yes	Yes	Varying hues to represent data age (Schlaisich *et al.*, 2004)
Orientation	Yes	Yes	Use of glyphs (Pang, 2001)
Shape	Yes	Yes	Tissot indicatrix (Slocum *et al*, 2005)
Color saturation	Yes	Yes	Color saturation of individual pixels to indicate accuracy (Schlaisich *et al.*, 2004)
Contour crispness	Yes	Yes	Line fuzziness indicating uncertainty of boundary (Slocum *et al.*, 2005),
Fill clarity\ Resolution	Yes	Yes	Fuzzy classification of imagery data (MacEachren *et al.*, 2005), Rain/snow and fog/haze techniques (Zuk, 2008)
Transparency	Yes	Yes	Transparency of uncertain 3D artifacts (Zuk, 2008),
Spacing (texture)	Yes	Yes	Spacing between iso lines to show weather model uncertainty (MacEachren *et al.*, 2005)
Perspective height	Yes	No	Using a 3D graph to show the quality attributes of an area (e.g. resolution) (Schlaisich *et al.*, 2004)

Previous methods have not been put into widespread use or implemented in commercial applications for a number of reasons. First, many of these visualization methods require a detailed assessment of the quality of the data, which is often not available. 3D objects stored in on-line warehouses such as Google Warehouse can be different representations of the same reality (see Figure 3.1). However, many of these objects do not come with any quality information and are hard to assess in terms of quality. One exception is the Google 3D Warehouse which does provide a five star rating and feedback system for individual models. However, this information is not provided or accessible once the model is visualized in the VG. Second, these methods are often very complex in nature, and hence difficult for non-expert geospatial data users to understand. For example, the use of glyphs to convey the direction and magnitude of uncertainty (see Figure 3.2) may not be very easy for a user with no background or formal training in geospatial data or statistics to comprehend. Finally, many of these methods would not be easily implemented on a large dataset such as all of the 3D objects in the Google Warehouse. A method such as using transparency on a 3D object to communicate the specific areas of the object which are more certain than others may be a very viable method by which to communicate uncertainty (see Figure 3.2), but modifying and assessing large amounts of 3D objects individually to use this method may not be feasible or effective.

Figure 3.1: An example of different representations of the same reality via models of the White House in the Google Warehouse

Figure 3.2: Use of transparency (left) (Zuk, 2008), and vector field of glyphs (right) (Wittenbrink *et al.*, 1996) to convey geospatial uncertainty

3.3 Categorizing Virtual Globe Users and Usages

Exploring the recent literature on VG users and usages shows that there is still much work to be done in building this domain of knowledge. Few studies characterize the users of VGs and for what VGs are being used. Grossner *et al.* (2008) identifies three categories of users: (1) non-expert users, such as a young child visiting a digital museum, (2) collaborative scientists and (3) current GIS users, such as governments that are utilizing GIS capabilities. To respond to this still limited understanding of VG users, Figure 3.3 presents a new classification, organized along two dimensions: the level of expertise of the user and the complexity of the task at hand.

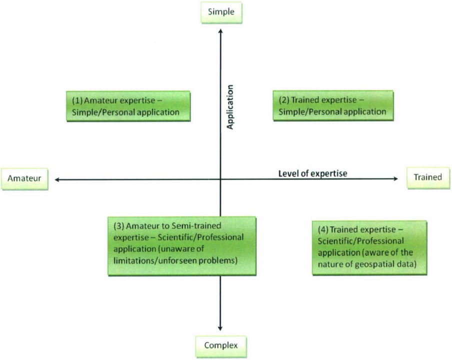

Figure 3.3: A classification of virtual globe users

Through doing this, four classes of VG users were defined:

(1) Amateur expertise – Simple/Personal application (e.g. viewing the road you live on using Google Earth's street view).

(2) Trained expertise – Simple/Personal application (e.g. navigating the streets of New York at eye level using Google Earth (Jones, 2007)).

(3) Amateur to intermediate expertise – Scientific/Professional application (e.g. the use of landscape visualizations from Google Earth by a Colorado group to visualize the skyline of the city (Sheppard & Cizek, 2009)).

(4) Trained expertise – Scientific/Professional application (e.g. dissemination of object-based change detection research results by researchers via Google Earth (Tiede & Lang, 2007)).

With most users of VGs being non-experts, it is important to design a new approach for visually communicating the geospatial data quality of the 3D objects in VGs. There is now a need for data quality communication to all types of users as existing data quality communication methods are too complex and will not be understandable by most non-expert users. This poses a need for novel methods of geospatial data quality communication.

The uses of VGs range from very simple to complex, including diverse analytical tasks (Goodchild, 2008; Grossner *et al.*, 2008; Sheppard & Cizek, 2009). Simple tasks can be viewing one's house, finding the location of a hotel, or getting driving directions. There are several examples of usages of VGs which are more complex. Some of the possible usages of VGs were for analytical decision making, disseminating research to the public and speeding up responses to natural disasters (Editorial, 2006). Tiede and Lang (2007) describe one study where 3D symbols representing the distribution of dwellings in refugee camps in an area of Tanzania were created using Google Sketchup. Sheppard and Cizek (2009) also exemplify several cases where VGs and data from VGs are used for analytical purposes. Two examples he gives are the use of Google Earth to design visualizations of a projected development in Colorado Springs, and a landscape visualization of a future possible flooding scenario in the Lower Mainland of BC, Canada, done by the Sierra Club. Such examples confirm the diversity of usages made of VGs. Since this research aims to visualize the geospatial data quality for 3D objects in VGs, the most important usages to be aware of are those which are complex and involve decision making. Understanding the quality of the geospatial data can then help to reduce undesirable results.

3.4 Visualization Approach

3.4.1 Introducing the Concept of Perceived Quality

Perceived Quality, of geospatial data is a novel concept which results from the consideration of the type of SDQ one would be communicating to users in a context where data quality is rarely formally assessed by data producers. In contrast to more traditional descriptions of SDQ, Perceived Quality is an expression of the perception which a user has about the quality of data. It is different from the concept of *internal quality* which measures, often quantitatively, the intrinsic quality of a dataset by comparing the data to a more accurate dataset. It is also different from the concept of *external quality* (i.e. fitness for use) as it does not assess how well a dataset fits the user requirements rather; it is assessing how the user perceives the quality of the data. In examining a model and looking at other users' commentary and feedback, the user gauges what the quality of the dataset.

3.4.2 Adapting a Symbiotic Approach

The *symbiotic approach* is a term referring to seeking a balance between the level of detail provided by a method and the simplicity of its implementation which considers three elements of the method; the usability by non-expert users, the technical feasibility, and the use of visualization elements such as visual variables. This trade-off is seen as necessary to reach solutions that can be widely adapted by a user community. By analogy with the biological meaning of "symbiotic", this approach assumes a mutual benefit to the users, developers and theoreticians of the method. The approach considers several issues, such as the richness of visualization vs. the ease of understanding, and also considers the technical feasibility of the visualization method.

Richness of visualization vs. ease of understanding

There are numerous examples of uncertainty visualization methods in the literature (e.g. Zuk, 2008; MacEachren *et al.*, 2005; Slocum *et al.*, 2005; Schlaisich *et al.*, 2004; Davis & Keller, 1997). These examples use techniques such as glyphs, wireframe, transparency, and various visual variables for visualizing geospatial data uncertainty or quality.

These methods for communicating geospatial data uncertainty may be very effective and are generally visually rich methods of communication. This being said, the ease of the general public's understanding must be weighed against the complexity and richness of the communication, particularly when the uncertainty communication is aimed towards non-expert users. Much of the general public may have trouble understanding what is represented in visualizations using complex methods, such as glyphs, to represent some type of uncertainty.

Even when the methods are less visually rich or complex, the question of whether a non-expert would understand a quality indicator being visualized should be considered. For example, the use of transparency to visualize the structural uncertainty in a 3D building may be easily understood by someone looking for this uncertainty, or someone who understands the nature of uncertainty, but may be too specific or difficult to comprehend for a non-expert user. This relates back to the argument that visualizing uncertainty may be far more effective for non-experts rather than methods such as statistical output (Buttenfield, 2003; Drecki, 2007). In the case of this research, it is the non-experts ability to understand the underlying richness and complexity of the visualization itself which needs to be reconsidered. Another related argument for simplicity is put forward by Bosworth where, in discussing the use of standards, he states that "the odds of failure are at least the square of the degrees of complexity of the standard" (Talking to DC, 2009).

Using the argument of simplicity in making a trade-off between visual richness and ease of understanding, one good way to explore the best methods for visualizing geospatial uncertainty in VGs is to examine what non-expert users do understand. Many popular websites use visual methods of communication of product quality, seller quality, etc. These popular websites often communicate the quality of some aspect of their websites, and this quality is user driven (e.g., feedback and ratings of products on amazon.com). User review systems are increasingly used and seen as important tools for on-line retail websites (Kim *et al.*, 2006; Qu *et al.*, 2008). Given that the users of these popular websites have a prior understanding of these systems, it can be assumed that they will most likely better understand the uncertainty in geospatial data which they are using if it is presented and/or user driven in a similar fashion.

Technical feasibility

No known examples of previous research in geospatial data uncertainty visualization have been implemented in popular geospatial applications. Considering the feasibility and usability of these communication methods may increase the chances of having the methods used. Many of the methods which are present in the research are undoubtedly effective ways in which to communicate uncertainty in geospatial data. One trade off to consider though, is whether it is realistic to implement the method in a widespread manner or on a large dataset. This is the trade-off between technical feasibility and complexity. Methods which use or change the dataset itself may be very difficult to implement on a large dataset. For example using techniques such as wireframe or transparency to communicate uncertainty in 3D objects could be very difficult and/or time consuming to apply to all of the 3D building models in Google Earth. Examining ways in which other domains currently visually communicate uncertainty and quality measures could therefore be a valid approach to determining which techniques are technically feasible.

3.4.3 Selecting the Five-Star Method

The exploration of methods used by other fields for visualizing the quality and reliability of data/products revealed that most popular websites use a five star rating method in conjunction with user feedback and individual ratings. Examples of such websites can be seen in Figure 3.4. Such consumer feedback methods are increasingly used for e-commerce and other websites (Kim *et al*., 2006; Qu *et al*., 2008; Thoms *et al*., 2010). They help consumers assess the quality of products before doing online transactions and build consumer trust in the website (Qu *et al*., 2008). The lack of existing metadata that could describe the quality of the objects in VGs did not allow the use of a number of traditional uncertainty visualization methods. The *Five Star Method* was chosen as this method is not subject to either of the constraints seen in implementing traditional visualization methods and is by far the most popular method used on the Internet to visualize the quality of products and sellers that have been assessed by users. This method fits with the symbiotic approach for several reasons. First, the method is one which is already implemented by a number of websites for communicating the quality of non-geospatial data. Second, most Internet users are familiar with this method, which makes it easily understood by a users having various level of expertise. Finally, the method provides a simple visualization, but also provides more detailed feedback through user comments and individual ratings, so the information communicated is still quite rich.

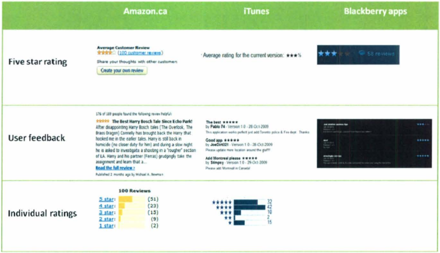

Figure 3.4: Example of the five star method from three popular websites

One common criticism about methods which rely on the knowledge of the crowd, or *crowd-sourcing*, is that they are not as reliable as knowledge generated by an expert. However, Haklay (2010), through a comparative study using OpenStreetMap, concludes that the quality of VGI datasets is not unlike datasets produced by professionals.

As these representations of the five star method in popular websites are not specific to geospatial data, a large part of this research project was assessing appropriate ways to geospatially visualize stars. This presented several challenges as to which visual variables and types of representation would be most appropriate to be visualized within a VG environment and understandable for end users.

3.4.4 Determining Representations of the Stars

Several possible ways that can be used to represent the *Five Star Method* were taken into consideration, three examples of which can be seen in Figure 3.5. This method is one which uses extrinsic and on-screen visualizations, wherein the visualization is part of the map environment and added to the map instead of being an integral part of the dataset itself. Therefore, this research had to consider the best way to transpose the *Five Star Method* to geospatial 3D objects in VGs. Several issues had to be taken into consideration for this. For example, issues such as where to place the stars in space with regards to the 3D object, which visual variables should be used and how should they be used, how the method could work with users changing their point of view in the VG, and how to deal with perspective and occlusion effects of 3D VGs.

Several of the visual variables were eliminated as possibilities for use in the five star visualization method. The use of size was eliminated due to the logical fact that the size of the symbol would change with the zoom distance in a VG environment. Also, colour hue was eliminated as it would be difficult to see some colours in contrast to all of the colours in the environment. The use of a colour scale was considered, but it was realized that often times the lightest saturation was hard to distinguish from the 3D object (building top).

Figure 3.5: Tests of the five star method using the visual variables size, number and colour

The next consideration for representing the five star method in a geospatial environment was whether to use 2D or 3D geometry for the symbology. As a 3D representation of the symbol is also a static object like the building itself, the symbol does not rotate with the user's view when changing the viewing angle in the VG. Therefore a 2D representation was chosen to be used as it can be viewed from any angle.

3.5 Perceived Quality

When considering the five star method the type of geospatial data quality being communicated is not one which is measured, such as error or positional accuracy, but rather it is a type of external quality which we have termed *Perceived Quality*. The Perceived Quality of geospatial data is the perception which a user has about the quality of data for their own needs. From examining a dataset (e.g. a 3D building model) and the comments and ratings which other users have made pertaining to this dataset, the user then gauges what the quality of the dataset is for their personal use (e.g. the dataset is good to use for assessing the aesthetic value of a real estate property) based on other users commentary and feedback.

This Perceived Quality can indeed include types of measured quality within the user comments, but the information in this feedback depends on the user's knowledge of the object in real life, and is also reflective of the user's purpose for viewing or utilizing the 3D object.

Using illustrations and describing Perceived Quality, we now will discuss the concept and how it works with the five star method:

1. User 1 creates a model of 3D object using an online application such as Google SketchUp. This model is then uploaded and stored in an on-line warehouse such as the Google Warehouse where other users can access and view the 3D object in a VG.
2. The model created by user 1 is viewed within a VG by users 2…n. These users proceed to assess the model by giving an overall rating (1-5 stars) to the model and by giving commentary through a user feedback box. The rating given and feedback may be general, about their real life knowledge of the representation, or may relate to why or why not the model fits their own purpose for using the model (Figure 3.6).
3. The accumulation of several ratings and comments for the model then allows for an overall rating of the model's Perceived Quality to be assigned (Figure 3.6).
4. Other users then access the model and view the overall rating, and can access the commentary and ratings provided by previous users. By filtering through the commentary and taking into consideration only the comments which pertain to

their own personal use of the representation, the users then formulate their own perceived quality of the 3D representation and decide whether it has a good fitness for use for them (Figure 3.7).

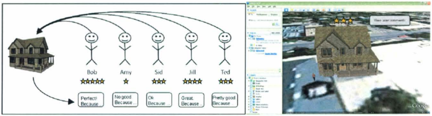

Figure 3.6: Perceived Quality evaluation (left) and average rating (right)

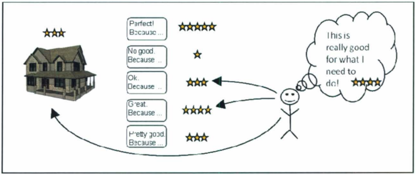

Figure 3.7: End users perception of the Perceived Quality

3.6 Prototype Development

The approach described in section 3.5 was prototyped into the software Google Earth. First, the sheer popularity of Google Earth was a large consideration, as the prototypes were to be tested by users that should be familiar with the technology. This way the majority of users would have used the application prior to the survey. Second, the technical aspects of creating prototypes using Google Earth was taken into consideration, as creating and implementing KML files is relatively straightforward.

Approximately 100 3D buildings in each of two cities were chosen for the prototypes. These 3D buildings were pre-existing in Google Earth. As the purpose of the prototypes was to determine which of the five star methods was perceived as the most effective by the users for visualization within a VG environment, the actual quality of the 3D buildings was not relevant. Each of the 3D buildings was assigned a random quality rating of one through five using a random number generator.

The two cities chosen for the prototypes were San Francisco and Prague (see Figure 3.8), because of their very different urban characteristics. By doing this, users would be able to experience the different representations in settings that posed unique problems. In Prague many of the buildings are at a very similar height, but are very close to one another. In downtown San Francisco, the buildings have greatly varying heights, and issues such as occlusion can be seen.

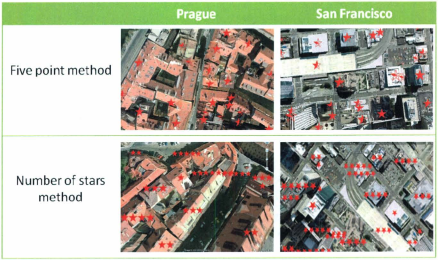

Figure 3.8: Examples of prototype of Five Point, and Number of Stars methods of visualization in San Francisco and Prague

The prototypes were made by creating KML files for each visualization method and city. KML files were created from a spreadsheet containing the altitude, latitude and longitude information gathered for each building from Google Earth, along with the randomly assigned rating. The four representations of the five star method that were chosen for the prototypes were the five point, number of stars, layered star, and number within star methods (Figure 3.9)

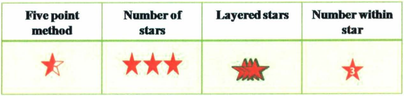

Figure 3.9: Four visualization methods used for prototypes

3.7 User Evaluation

A user survey was conducted with both expert and non-expert users to evaluate the five star approach developed. The survey was performed by 40 users in total, including 20 expert and 20 non-expert users of geospatial data, all of whom had used VGs before.

Their level of knowledge of GIS and VGs was first assessed through questions at the start of the survey. Then, participants were asked to explore the 3D data using both prototypes (San Francisco and Prague) with each of the five star methods. To avoid any bias that could be introduced through the order in which each of the visualization methods were viewed, each participant was given a different ordering determined using the Latin Square method which gave each participant a different order in which to view the four Five Star Methods.

After viewing each of the four methods in both cities, the participants were then asked to rank the four methods in relation to their ease of understanding (one being the easiest to understand). They were then asked to provide two reasons each for why they chose the first and last methods.

The results of this ranking were then statistically analysed in order to answer the following questions:

(1) Is there any significant difference in the rankings of the five star methods between survey participants?

(2) Is there any difference between the rankings of the five star methods between expert and non-expert GIS users?

Answering these two questions helps answering the last research question of this study (see section 1.4) which is "Which of the visualization method(s) implemented in the visualization prototypes is most effective for communicating the geospatial data quality to users."

Non-parametric statistics were chosen for these analyses because of the low sample size (n=40 with 20 persons per group). Warner (2007) suggests that non-parametric tests should be considered when $n < 20$ or definitely used when $n < 10$ per group. SPSS PASW Statistics 18 was used to perform the statistical analyses. Friedman's Test was used to answer Question 1 by comparing the mean ranks of each of the four methods. Results indicated that there was a significant difference in the rankings ($p<0.000$). Results indicated that the method 'number within star' was consistently ranked higher, and the *number of stars* method was consistently ranked lower. It should also be noted that 45% of the participants ranked the *five point* method second and 30% ranked it first, making it the second most favoured method.

The Mann-Whitney Test was used to answer Question 2. By comparing the two populations (expert vs. non-expert) it was found that there was no significant difference between the rankings of experts vs. non-experts (Exact Significance [2*(1-tailed Sig.)]: $0.192 < p < 0.602$). This result indicates that the level of expertise seems to have no influence on the ranking of the methods done by the users.

Qualitative data from written feedback about why each user ranked a method as first or last gave further insight as to why these methods were selected. The main reasoning for choosing the *number within star* method as number one was that it was the easiest to understand as it used a number, and that it was not as cluttered and crowded as other methods. The primary reasons for placing the *number of stars* method fourth was that it was too cluttered and crowded which obstructed the 3D buildings and made the rating confusing to read. These results are interesting as they contrast with what is known from other environments where five stars are often used to visualize quality. Therefore, when considering applying such a method in a 3D environment it is certainly important to realize that it may be perceived differently by users.

3.8 Conclusion

Visualizing SDQ has been the focus of a number of research projects in the past, but the proposed methods, while conceptually interesting, have not been put into mainstream or widespread commercial use. This is partly because these methods can be challenging to implement and require measurements of data quality. Many of the methods are also not easily understandable by non-expert users of geospatial data, and it is now these users who mostly use VG systems. This paper has presented a new classification of VG users. It also provides a novel way to collect and display quality of 3D objects in VGs which is based on a symbiotic approach, which combines a detailed volunteered assessment (feedback) with a simplified display (stars). The five star method communicates *perceived quality*. The visualization approach in this research was tested and developed using 3D buildings in Google Earth. Of the four five star prototypes developed, it was determined that using a number within a star to visually communicate the quality ranking of a 3D object in a VG is the easiest for users to understand because of the method's visual simplicity and less clutter and confusion compared to the other methods. Such an approach can be used for any type of 3D object in any VG.

The visualization approach proposed in this project is dependent on multiple users having knowledge and perceptions of a 3D object. Consequently, the method of visualization may have limitations in certain geographical areas. Future works for this research would be to develop the visualization approach further as an add-on application, or to explore the use of filters or sliders in displaying ratings.

3.9 References

Beard, K., Mackaness, W., 1993. Visual Access to Data Quality in Geographic Information Systems. *Cartographica*, 30:2&3, pp. 37-47

Butler, D., 2006. The web-wide world. *Nature*, 439:7078, pp. 776-778

Buttenfield, B. P., 1993. Representing Data Quality. *Cartographica*, 30:2&3, pp. 1-7

Chrisman, N. R., 1983. The role of quality information in the long-term functioning of a geographic information system. In: Douglas D H (eds.), *Proceedings of AUTO-CARTO 6*. Ottawa, ON, Canadian Institute of Surveying and Canadian Cartographic Association: 303–21

Coleman, D. J., 2010. The Potential and Early Limitations of Volunteered Geographic Information. *Geomatica* 64:2, pp. 209-219

Coleman, D. J., Georgiadou, Y., Labonte, J., 2009. Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*. 4, pp. 332-358

Davis, P., Keller, C., 1997. Modeling and Visualizing Multiple Spatial Uncertainties. *Computers & Geosciences*. 23:4, pp. 397-408

Devillers, R., Jeansoulin, R., 2006. Spatial Data Quality: Concepts. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 31-42

Drecki, I., 2002. Visualization of Uncertainty in Geographical Data. In: Shi, W., Fisher, P. F., Goodchild, M. F., (Eds.), *Spatial Data Quality*. Taylor & Francis, New York: NY, pp. 140-160

Editorial. 2006. Think global. *Nature*, 439:7078, pp. 763

FGDC (Federal Geographic Data Committee), 1991. *Spatial Data Transfer Standard*. Washington, DC.: Department of the Interior

Girres, J., Touya, G., 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14:4, pp. 435-459

Goodchild, M. F., 2008. Spatial Accuracy 2.0. In: J.-X. Zhang and M. F. Goodchild (Eds.). Spatial Uncertainty. *Proceeding of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*. Volume 1. Liverpool: World Academic Union, pp. 1–7

Grossner, K. E., Goodchild, M. F., Clarke, K. C., 2008. Defining a Digital Earth System. *Transactions in GIS*, 12:1, pp. 145-160

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*. 37, pp. 682-703

Haklay, M., Singleton, A., Parker, C., 2008. Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*. 2:6, pp. 2011-2039

Hoeber, O., Wilson, G., Harding, S., Enguehard, R., Devillers, R., 2010. Visually representing geo-temporal differences. *IEEE Conference on Visual Analytics Science and Technology*, Salt Lake City, USA, 2010

Hunter, G. J., Goodchild, M. F., 1996. Communicating Uncertainty in Spatial Databases. *Transactions in GIS*, 1:1, pp. 13-24

Jones, M., 2007. Google's Geospatial Organizing Principle. *IEEE Computer Graphics and Applications*. 27: 4, pp. 8-13

Kim, S., Pantel, P., Chklovski, T., Pennacchiotti, M., 2006. Automatically Assessing Review Helpfulness. In *EMNLP 2006: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Sydney, pp. 423–430

Kresse, W., Fadaie, K., 2004. *ISO Standards for Geographic Information*. Springer, Germany

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, S., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32:3, pp. 139-160

Pang, A., 2001. Visualizing Uncertainty in Geo-spatial Data. In: *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*. pp. 1-14

Peter, M., 2009. Presentation and Evaluation of Inconsistencies in Multiply Represented 3D Building Models. In: Kurt Rothermel, Dieter Fritsch, Wolfgang Blochinger, Frank Dürr (eds.), *Quality of Context, First International Workshop, QuaCon 2009, Stuttgart, Germany, June 25-26, 2009*. Springer, pp. 156-163

Qu, Z., Zhang, H., Li, H., 2008. Determinants of online merchant rating: Content analysis of consumer comments about Yahoo merchants. *Decision Support Systems* 46, pp. 440-449

Rakshit, R., Ogneva-Himmelberger, Y., 2008. Application of Virtual Globes in Education. *Geography Compass*. 2:6, pp. 1995-2010

Schlaisich, I., Mountrakis, G., Agouris, P., 2004. Visualization of Image Quality in Distributed Spatial Databases. In: M. O. Altan (Editor), *International Society for Photogrammetry and Remote Sensing XXth Congress*. Istanbul, Turkey

Sheppard, S. R. G., Cizek, P., 2009. The ethics of Google Earth: Crossing thresholds from spatial data to landscape visualisation. *Journal of Environmental Management*. 90, pp. 2102-2117

Slocum, T., McMaster, R., Kessler, F., Howard, H., 2005. *Thematic Cartography and Geographic Visualization*. Pearson Prentice Hall, NJ, 518 p

Talking to DC. In Adam Bosworth's Weblog. Last accessed, June 2011. <http://adambosworth.net/2009/10/29/talking-to-dc/>

Thoms, B., Garrett, N., Ryan, T., 2010. The Design and Evaluation of a Peer Ratings System for Online Learning Communities. In *HICSS 43: Proceedings of the 43rd Hawaii International Conference on System Sciences*, Koloa, Kauai, Hawaii, USA, IEEE, 10 p

Tiede, D., Lang, S., 2007. Analytical 3D Views and Virtual Globes – Putting Analytical Results into Spatial Context. *Paper from: ISPRS Joint Workshop "Visualization and Exploration of Geospatial Data"*. 6 p

Tuttle, B. T., Anderson, S., Huff, R., 2008. Virtual Globes: An Overview of Their History, Uses, and Future Challenges. *Geography Compass*. 2:5, pp. 1478-1505

Warner, M., 2008. *Applied statistics: from bivariate through multivariate techniques*. SAGE Publications, Inc, CA

Wittenbrink, C. M., Pang, A. T., Lodha, S. K., 1996. Glyphs for visualizing uncertainty in vector fields. In the *IEEE Transactions on Visualization and Computer Graphics*. 2:3, pp. 266-279

Zuk, T. D., 2008. Visualizing Uncertainty. *PhD thesis*, Department of Computer Science, Calgary; Canada, 336 p

Chapter 4: Expanded Results

4.1 Introduction

This thesis presents a novel method for the visual communication of the geospatial data quality of 3D objects in VGs. This chapter aims to present in further detail some of the steps from the research process that were not covered in the Chapter 3 because of space constraints due to the journal requirements in which the paper is to be published. First we will discuss the general approach taken in this thesis, followed by a more detailed description of the concept of perceived quality. Finally, we will present different aspects of the prototype development and the results from the user evaluation.

4.2 Approach

The general approach taken for the visualization method is one which we have termed a *symbiotic approach*, which is a trade-off between the technical feasibility of the method, its ease of understanding for the users, and the richness of the visualization. The approach was discussed in detail in section 3.4.2. Adapting the five-star system allowed to achieve the objective of designing an approach for visualizing the quality of 3D geospatial data within VGs.

Taking into account the need for a symbiotic approach, the lack of widespread implementation of previously proposed methods, and the knowledge that much of the users of VGs are users with no formal expertise in geospatial data (see section 3.4.3), methods for visually communicating the quality of products and services were researched from other fields. Implementing a five star rating method to display the overall quality of the 3D object simplifies the visualization and improves the ease of understanding. Then, by supplying additional information, such as individual user comments on the quality and individual ratings, the user then also gets a detailed assessment that they can use to assess the quality along with the simplified display. Examples of all of these elements of the communication of quality can be seen in popular websites such as Amazon, iTunes, and eBay (see Figure 4.1). Examining these other methods used to visualize data quality helped in achieving part of the second objective by identifying and classifying methods by which quality more generally can be visualized.

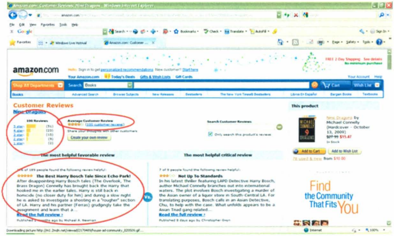

Figure 4.1: Example of five star rating and user feedback method from amazon.com

4.3 Perceived Quality

The type of geospatial data quality being communicated through this five star rating system has not been measured quantitatively, but is rather a form of external quality, which we have termed *perceived quality*. This is the perception that users have about the quality of data for their own needs and it is formulated through examining other users' comments and feedback. To illustrate this concept, Figure 4.2 shows perceived quality broken down into four simple steps;

(1) A user (named Rae) creates a model of a 3D object using an online application such as Google SketchUp.

(2) The model created by Rae is viewed within a VG by several other users. These users assess the quality of the model by giving an overall rating (1-5 stars) to the model and by giving commentary through a user feedback box. The rating given and the feedback may be general, or may relate to why or why not the model fits their own purposes.

(3) The accumulation of several ratings and comments on the model then allows for an overall rating for the model's external quality to be calculated.

(4) Another user (Dan) then accesses the model and views the overall rating and the individual user's ratings and commentaries. By browsing through the commentary and selecting the ones that pertain to their own personal use of the representation, the user then formulates their own perceived quality of the 3D representation.

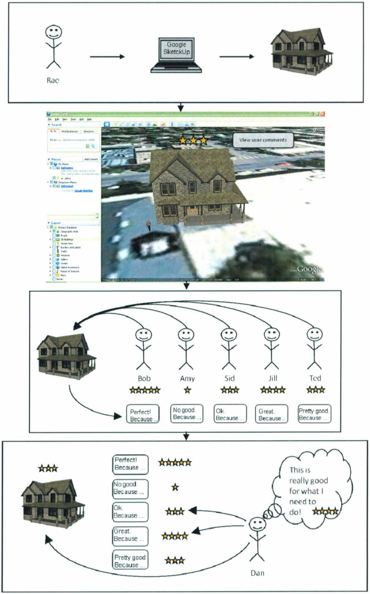

Figure 4.2: Illustration of the four steps for explaining Perceived Quality

4.4 Prototypes and User Evaluation

The next step in the research process was to implement software prototypes using the five star method. This step helped achieve the fourth research objective, being to implement and test the approach in a visual prototype. Prototypes were created in Google Earth using KML files. Many variations of the five star method were considered using different geometries and variations of the visual variables (see section 3.4.4). Ultimately four representations of the five star method were chosen for use in the prototypes (see Figure 4.3). These four representations were applied to approximately 100 3D buildings in both San Francisco, USA, and Prague, Czech Republic, in Google Earth. The method by which the KML files were created, and the reasoning for choosing these cities can be seen in section 3.6.

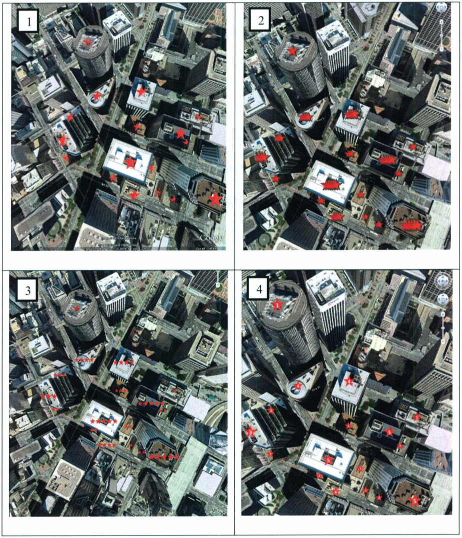

Figure 4.3: Images of prototypes for five point (1), layered (2), number of stars (3), and number within stars methods (4) in San Francisco

Following the creation of the prototypes a user survey was conducted to assess which of the five star visualization methods was best for the users to understand (see Appendix A). The survey was administered to 40 participants (20 non-experts and 20 experts in geospatial data). The sample size was defined to allow significant statistical tests to be performed. Participants were asked to navigate through the prototypes and rank each of the four methods based on how easy the visualization methods were to understand.

Friedman's statistical test compared the mean ranks of each of the four methods. The result was statistically significant ($p < .0001$) and showed that the method *number within star* was consistently ranked higher, and the *number of stars* method was consistently ranked lower. A more complete picture of how the methods were ranked by participants can be seen in Table 4.1. Subsequently, a Mann-Whitney test was used to compare experts with non-experts. It was found that there was no significant difference between the rankings of these two populations (Exact Significance [2*(1-tailed Sig.)] $0.192 < p < 0.602$), meaning that there was an overall consensus for the best and worst of the four methods regardless of level of expertise.

Table 4.1: Percentages of survey participants ranking of each of the five star methods (n=40)

Rank	Five point method	Number of stars	Layered stars	Number within star
Rank 1	30%	2.5%	5%	**62.5%**
Rank 2	**45%**	10%	20%	25%
Rank 3	15%	25%	**47.5%**	12.5%
Rank 4	10%	**62.5%**	27.5%	0%

Qualitative data were also collected to understand more in depth the reasons for which survey participants ranked the number within star method first, and the number of stars method last. Participants were asked to provide two reasons for each of their first and last ranking selections. This participant feedback can be seen in Tables 4.2 and 4.3. Additionally, a summary of the user comments for the top methods ranked first (method within star) and last (number of stars) is presented. The main reasons for which participants chose *number within star* for their first ranked method are: there is less clutter on the screen, it is generally the easiest to understand, it is aesthetically pleasing, and the use of numbers is easier to quickly comprehend the rating (Figure 4.4). The major reasons for participants choosing *number of stars* as their method ranked last are: it is confusing to look at, there is too much clutter and crowding on the screen, it is not aesthetically pleasing, the symbol size is obstructive of the buildings, and it is generally complicated (Figure 4.5).

Table 4.2: Qualitative user feedback on why they ranked one method first or easiest to understand

Method ranked first	User feedback
Number within star	
Non-Expert	- Because it has a visual (numbers) for people to see more clearly - Because it's not cluttered like the other stars. It's simple to distinguish
Expert	- Immediately understandable - Very clean/not as cluttered as the other methods
Non-Expert	- Tells you the actual numbers - Didn't cover up buildings too much
Non-Expert	- Gives simple numbers

	- Easy to understand
Non-Expert	- It looked good even when zoomed out - It was not cluttered
Non-Expert	- It was the easiest to understand - It was the neatest looking
Non-Expert	- Less crowding of information - Immediate information (no need for counting)
Non-Expert	- Numbers in the stars clarify the ratings better - Use of one star instead of many makes it easier to read
Non-Expert	- Easy to read, easy to look at - Ratings don't overlap one another
Non-Expert	- The numbers within the stars makes it easy and fast to see the rating - One star one each building is less crowded and easier to see the buildings. Also, less overwhelming
Non-Expert	- Clear and easy to see - Visually neater than the rest
Non-Expert	- Very clear to see - Not bunched up
Non-Expert	- It was much clearer to see - Looked tidier than other methods
Non-Expert	- It was easier to count the stars - There is less stars on the map so it's less congested
Non-Expert	- Easy to see when you zoomed in and out. You could see the quality even at lower resolution - Did not have to count stars
Expert	- Did not have to think about it, self evident, clear - No overlap between symbols
Expert	- Can tell which stars belong to which buildings and there is little overlap of stars that belong to the same buildings - It is obvious at first glance what the rating is
Expert	- Simplicity of understanding
Expert	- Clarity of the symbol - # ranks the building from worst (1) to best (5) - No overlap between symbols for different buildings
Expert	- There is a single image, less confusing, easier to look at - The ranking is in the image, seems easier to understand
Expert	- It is not cluttered - It is obvious and doesn't need interpretation
Expert	- Easiest to read at a glance - Shows up well on all backgrounds
Expert	- Can easily identify the number when zoomed in or out - Least cluttered and easy to see
Expert	- All other methods had instances when it could be hard to make out the ranking
Expert	- Faster recognition of its value - More organized and visually appealing
Expert	- The number system is more clear - Less clutter allows you to concentrate on what you are viewing especially when there are a multitude of rates in the same area

Five point method	
Non-Expert	- Gives you a clearer picture of the building seeing more of the building - Easy to understand five points one star on each building
Non-Expert	- Least confusing/easy to grasp - Does not overcrowd screen
Non-Expert	- Took up less space on the map - Very clear to see
Non-Expert	- Clarity - Less on screen
Expert	- It is easy to think of empty (no color) part of the star as "nothing" or "not counted" - Easy to see one star per building
Expert	- Easy to attach star to building - Easy to understand which is best/worst buildings
Expert	- Visually intuitive and clear - Does not block the view – not as overwhelmingly as the other methods
Expert	- Ease of understanding - Level of clutter
Expert	- Easy to interpret the rating - Easy to associate which rating belonged to which building
Expert	- More concise symbol - The more the star is filled, the better the quality – easiest to interpret at a glance
Expert	- It is the two top ranked choices appear less cluttered and easier to associate w/ correct buildings - Quicker to read the five-star ranking
Layered stars	
Expert	- Reduced clutter compared to 2-number - Still quickly gives indication of magnitude compared to 1 five pt. and 4 number within
Non-Expert	- The stars are close together so they appear neatly; they don't overlay other buildings ratings - Convenient method of star ratings
Number of Stars	
Non-Expert	- Easiest to see at a glance

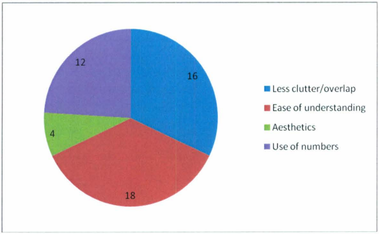

Figure 4.4: Summary of comments for method most often ranked first – number within star

Table 4.3: Qualitative survey user feedback on why they ranked one method fourth or hardest to understand

Method ranked last	User feedback
Number of stars	
Non-Expert	- Not clear which stars belong to which buildings - All over crowded picture. Making it difficult to get a clear picture
Non-Expert	- Too many stars makes it complicated - Hard to understand
Non-Expert	- Too many stars on the screen - Could not make out which building the stars belonged to when zoomed out
Non-Expert	- Too noisy, annoying to look at - Too much overlap between individual points
Non-Expert	- Too many stars overlapping - Hard to pick out which stars are with each building
Non-Expert	- Stars are too big cover too much of the building - Ratings start to run together, no definition between ratings
Non-Expert	- Difficult to tell which stars belong to which building - Too many red stars on screen, stars are dominant
Non-Expert	- It's hard to understand which stars are for each building - There are too many stars on the map
Non-Expert	- Looks sloppy and crowded - Overlapping. May be hard to count number of stars correctly
Non-Expert	- Too many stars, was confusing - Have to count too many stars
Non-Expert	- Layering of the stars was visually irritating -There were too many stars visible on the screen
Expert	- Too difficult to visualize - understand
Expert	- Very cluttered and hard to distinguish between close buildings - Takes up a lot of space on the screen
Expert	- Hard to not have overlap between symbols if there is a high density of 3D buildings - Hard to make out exactly which building is being rated
Expert	- Too much clutter, particularly bad when zoomed out and all methods exhibit some clutter
Expert	- Difficult to relate stars to building (overlap is confusing) - Overlapping stars (when next to each other) can make the stars of adjacent buildings seem like they may be for a single building
Expert	- Hard to tell which stars belong to which buildings - Stars are varied in size, the size does not seem to be related to the rating

Expert	- Block the view when zoomed out then overlap on other building (mainly for Prague) - Add too much visual noise to the view
Expert	- Too crowded/takes up too much room. Symbol tends to overlap multiple buildings - Overlap with ratings for other buildings
Expert	- Not good when buildings are closer to each other. Appears too crowded - Hard to tell the number of stars when zoomed out
Expert	- Stars appeared jumbled and overlapped - Harder to determine which stars belong to which building feature
Expert	- Clutter, very hard to tell if the stars are from another building - Hard to see the building in under the stars
Expert	-Too much clutter on the area obscures the subject matter - Not as easy to identify the number when they are layer. Miss-counting can occur when the data is viewed quickly
Expert	- Too confusing. Visually "a mess" - Hard to determine which rating belonged to which building
Expert	- Cluttered – Not visually appealing - Difficult to associate w/ current building in densely built up areas
Expert	- Too many symbols, at smaller scales they run together - At first, not easy to identify which building they reference. One must zoon in to do so
Five point method	
Non-Expert	- Covered up buildings - Had to count to read
Non-Expert	- Deceiving, one star is actually five star rating not one star rating - Sometimes hard to see the missing piece of the star
Non-Expert	- The appearance of the star shading differs according to the background color, not always easy to see - Not obvious what number of stars it represents. Takes more time
Non-Expert	- Hard to see
Layered stars	
Non-Expert	- It's very cluttered and you have to physically put your finger on the screen - I don't like it being layered – it looks mesmerizing
Non-Expert	- It's too confusing - It looks dumb
Non-Expert	- Stars were very jumbled together. Hard to see if 3 stars or 4 - Made the map look messy

Non-Expert	- Didn't like stacked stars
Non-Expert	- Was too confusing - Looks too cluttered
Expert	- Clutter - Harder to determine what building is being represented by the stars
Expert	- Hard to see the number of stars -Makes image look too busy and cluttered
Expert	- Must really look closely to see the rating compared to others - Kind of ugly I guess
Expert	- It is cluttered - It takes too much time to count the stars because they're close together
Expert	- Found it too hard to count - Visually hard to differentiate between 4 and 5

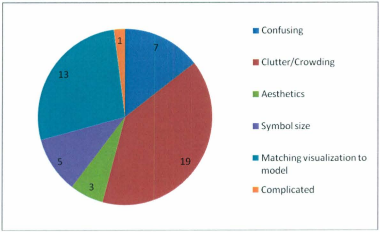

Figure 4.5: Summary of comments for method most often ranked fourth – number of stars

The user survey also included a section where participants viewed four 3D representations of the Empire State Building in Google Earth (see Appendix A), once with no quality visualization and once with it, and were asked to rank the quality of the buildings from best to worst in each scenario (with and without the five star rankings). The 3D building representations were uploaded to Google Earth from the Google Warehouse, and were created by various users. Each of the buildings had previously been assigned a quality ranking in the Google Warehouse, so four buildings with different qualities were chosen. The survey participants were then asked to comment on whether viewing the quality information the second time they ranked the quality of the objects affected their ranking. The purpose of this exercise was to see whether viewing the quality visualization along with the building helped or influenced the participants' decision of what the quality of the building was. No statistical tests were performed on the quantitative data collected from the ranking of the buildings, as 24 of the 40 participants ranked the quality of the buildings in the same order before and after they were presented with the quality visualization. This left too small of a sample size to compare each of the two sets of rankings and get meaningful results. The qualitative data however provided some insight into the participants' perception of how the addition of a visualization of the geospatial data quality for the 3D objects affected their rankings.

When asked "Did viewing the quality information about the 3D objects (buildings) affect your rankings of them?", 15 of the 40 survey participants simply stated no because their rankings did not change from part 1 to part 3, but majority of the other participants provided feedback confirming that the quality information was indeed useful to them (see Table 10).

Table 4.4: Table showing survey participant feedback pertaining to the affects of viewing quality information when ranking 3D objects quality

User	User feedback
Non-Expert	It didn't affect my ranking of them because I found my view different from the rankings provided based on my own opinion
Non-Expert	Yes, because before I saw the rankings I was confused about whether building A or D looked better
Non-Expert	Not really, I thought building B looked best but didn't pay attention to the stars above it. Same as C looks worst and was rated worst
Non-Expert	It did not, however A and B both looked similar in quality. If I had not seen the buildings in part one I may have been influenced by the star rating
Non-Expert	Yes, a bit. I couldn't help but notice the star ratings
Non-Expert	The ranking changed because knowing other peoples opinion influenced my decision on how I originally viewed the buildings
Non-Expert	Yes, quite a bit. I was unsure about what to look for to judge a buildings quality. Seeing the rankings provided me with a guideline. I think after a while users will establish a good sense of whether or not a building is modelled well or poorly
Non-Expert	Yes it did change my ranking because it helps you pick out more detail of the buildings
Non-Expert	The viewing of the rankings affected my ranking of building 1 and 2. Seeing the higher ranking of 2 caused me to examine the building more closely and notice the increased amount of detail
Non-Expert	The quality information confirmed (supported my ranking
Expert	Slightly. I needed to re-evaluate my decisions in part one because of the star rating
Expert	My ranking did not change, but I probably would have been influenced by the stars, especially for ranking the top two buildings which was a tossup the first time around
Expert	Not really, my eye was focused on the building image
Expert	Yes, it changed my opinion because I wasn't sure of the quality of three of them and knowing what other people rated them helped me decide
Expert	Slightly, but did not change my best and worst quality selection
Expert	It did, at a glance A and B were hard to judge but seeing the reviews swayed my opinion in a different direction
Expert	Yes. Forced me to consider other peoples evaluation of the buildings
Expert	The rankings were the same as in part one, but I felt more confident with my rankings after viewing the star ratings
Expert	The rating can be very subjective. It's hard to tell the difference between buildings without being given a ranking
Expert	It did not change my ranking, but it did make me look closer at A and D to see why I ranked them differently than the star system. In my opinion, D's windows look more realistic than A's! However, I know very little about the Empire State Building (i.e. I have no basis of comparison with reality)…
Expert	No, but it did instil confidence
Expert	No (But I think it would if it was a building that I had never seen or heard of before)
Expert	Did not for the best and worst quality but in some ways affected my decision for the middle one, the one for which the quality difference is less important
Expert	Yes, the quality information is based on knowledge from people who have a better

	understanding of what the building actually looks like. The quality information provides a measure of accuracy of the model based on other peoples opinion, thus it would influence my opinion on the model
Expert	Yes, I found it hard to rank between building A and D (in part 1). They both look great, but with the quality information available, it was relatively easy to decide

Performing this user evaluation and gathering both qualitative and quantitative data allowed us to better understand which of the five star methods was most preferred and why this was. The fact that all users completed the survey also helped validate that users could understand the quality of geospatial 3D models in a VG environment, and this also shows the efficiency of the method as all expert and non-expert users alike did not have to spend time learning or trying to understand the visualization method. Furthermore, the prototype itself being implemented in an existing VG environment shows the technical feasibility of the method itself. The coding and data collection for the prototype took approximately a week, and the execution of the user survey and subsequent analysis of the data took several weeks.

Chapter 5: Discussion and Conclusions

5.1 Summary

With an increasing use of VGs by a broad range of users, visually communicating the geospatial data quality of 3D objects VG is becoming increasingly important. Most of the methods proposed for communicating geospatial data quality or uncertainty can be difficult to implement, or are not always easy to understand by non-expert users. Other domains have successfully implemented methods for visually communicating the quality or reliability of products or data to lay users, typically using user feedback and rating systems. This approach was adapted for use on 3D data in a VG. This thesis designed, tested and validated an approach for communicating visually the perceived quality of 3D objects in VGs. It used a symbiotic approach which seeks for a balance between simplicity and detail of the visualization method. The visualization method developed communicates the quality of a 3D object as perceived by users examining the model, gauging the quality of the object in the light of their personal need and context. Several visualizations were tested and the top ranking method, *number within star*, was then identified using a user survey. These results imply then, that when applying such a method in a 3D environment, we need to consider using different approaches than those used in 2D environments (e.g. other environments where five separate stars are often used to visualize quality).

This thesis posed five specific research questions (presented in the section 1.4) that have been answered as follows:

- **What are the different types of users of virtual globes?**

VGs are a very recent technology. As a consequence, the review of the literature on VG users and usages revealed a generally incomplete understanding of the types of users and usages. To answer this first research question, a new classification of VG users was proposed, based on a review of the literature. The classification was structured around two axes: 1) the level of expertise in geospatial data of the user and 2) the level of complexity of the use made of the VG (see section 2.2.2). This resulted in identifying four types of VG users; (1) Amateur expertise – Simple/Personal application, (2) Trained expertise – Simple/Personal application, (3) Amateur to Semi-trained expertise – Scientific/Professional application, (4) Trained expertise – Scientific/Professional application.

- **What methods for geospatial data quality visualization have been previously explored?**

Knowledge of the breadth and types of existing methods proposed for visualizing geospatial data quality and uncertainty were identified through a review of the literature. These methods were classified according to 13 visual variables (see section 2.3.3). Identifying and classifying these methods served as an essential step in understanding what potential methods or aspects of methods could be implemented in visualizing geospatial data quality for 3D objects within a VG environment. Answering this question also gave insight into the limitations of these previously explored methods when considering the technical feasibility and level of expertise that would be needed for a user to understand the visualization.

- **Which factors are important to consider in the development of an approach for implementing geospatial data quality visualizations for 3D objects in a virtual globe?**

The review of the literature done on the existing visualization methods confirmed the diversity of methods that are available. The fact that none of them is used in current mainstream geospatial technologies indicates among other things the difficulty to implement such approach. This difficulty can result from a number of causes, such as the technical feasibility to implement the approach and the availability of quality or uncertainty assessments. As a consequence, two factors were identified for the development of the approach (see section 3.4.2): (1) the level of detail the method provides about data quality vs. the simplicity and ease of understanding of the method, where the method must communicate sufficient detail as to the geospatial data quality of the data, but must also be understandable to expert and non-expert users alike, (2) the technical feasibility of its implementation, where the method should not be too complex to be applied in a widespread manner to many different 3D models.

- **Which method or set of methods can be used to communicate geospatial data quality of 3D objects to non-expert users of virtual globes?**

Most methods proposed in the GIS literature for visualizing data quality or uncertainty can be difficult to implement in VGs due to the fact that they strongly rely on existing quantitative assessments of data quality. As a consequence, a broader search was done to see how other fields have successfully implemented methods for communicating the reliability or quality of data/products to lay users. Many methods rely on a *five star* rating system and user feedback, such as those used by websites such as Amazon, iTunes, and E-Bay. This thesis has been inspired by these methods which combine ease of representation (i.e. five stars) richness of the information (i.e. comments from users) and ease of implementation (i.e. no need for existing metadata and the use of simple symbols for the visualization). The method was adapted to a geographic visualization for 3D objects in the VGs environment (see sections 3.4.2 and 3.4.3) and has provided a novel way to collect and display quality of 3D objects in VGs. This method allows non-expert users to understand the quality of the data they are using for decision-making, and is a new approach which is based on the users' perception of quality (see section 3.5).

- **Which of the visualization method(s) implemented in the visualization prototypes best visually represents the geospatial data quality?**

Several ways for representing the five star method in a geographic space were considered for visualization prototypes. The use of 2D and 3D stars was tested. In addition, a number of visual variables have been experimented to display the information (see section 3.4.4). From these possible visualizations, a subset of four methods was selected for prototyping (see section 3.6). The four visualization methods were then tested through a user evaluation and then statistically validated to determine the most effective one (see section 3.7). It was found that the *number within star* method was consistently ranked highest, with the *five point* method ranking second best. The evaluation also presented no significant difference between the rankings of experts vs. non-experts, which reinforces that the *number within star* method was the overall consensus for the best of the four methods. Qualitative data obtained from this survey also gave insight as to why this method and the *five point* method were preferred (see section 4.4).

The research hypothesis (section 1.3) was that it was possible to visually communicate the geospatial data quality of 3D objects in VGs to non-expert users through a method which is both technically feasible and easily understandable by non-expert users. The selection of the five-star methods allowed having a method geared towards non-expert users, as this method has been successfully used in other fields (see section 3.4.3). This method was proved to be very easy to implement within a prototype (see section 3.6), by using a small selection of map symbols that can be displayed as KML files in the prototype. The user evaluation (see section 3.7) also confirmed that the approach was easy to understand by users, both experts and non-experts in geospatial information. We can then conclude that the research hypothesis has been confirmed by this study.

5.2 Limitations and opportunities

While the thesis has provided answers to several questions, there are however still a number of limitations to consider and opportunities to expand and build upon the research. These limitations and opportunities relate to the methodology, user survey, scope of the research, and technical aspects of the prototypes.

One limitation of this research can be seen in the user survey. As the survey was not the major goal of the research, the population sample was limited (20), and therefore the statistical analysis was limited to those tests for a small sample size. While the sample size used is statistically valid for the use of non-parametric tests (see section 3.7), a larger group of users would have been interesting to get, as no generalizations were made from the results of the survey. In this case there is then an opportunity for future works to include a larger collection of quantitative data and more complex statistical analysis, perhaps via an on-line survey.

The visualization method proposed in this project relies on the perception users have of the quality of 3D objects. While this is likely the only way to assess the quality of the data in absence of a formal quality assessment that would require ground-truthing data, it presents limitations similar to what is known from other types of volunteered geographic information (VGI). The reliability of the assessment is often a function of the number of users that have provided an assessment. Haklay (2010) describes a similar problem with OpenStreetMap data where the quality of the data increases in urban city areas but drops considerably in rural areas. This problem results from having a lower number of contributors in rural regions, which decreases the quality of the data (e.g. decreased completeness and accuracy).

Opportunities for future work also stem from this research. The scope of this project only required testing the concepts using a prototype (see section 3.6). Developing a fully functional system for the five star method with a VG would be interesting in order to further test the ideas. Also, future research into users' motivations and willingness to assess the *perceived quality* of 3D objects in VGs would be very interesting. Similar studies assessing users motivations and willingness to contribute VGI have already been conducted (Coleman *et al.*, 2010; Coleman, 2009; Budhathoki, 2010).

As part of the research, the use of a quality slider was also explored in order to select dynamically only objects that have a certain level of quality. The incorporation of this into a working system, and the exploration of other features to enhance the system may also be an interesting way to further the research as a future work.

Finally, this research provides a novel way to collect and display quality of 3D objects in VGs, and introduces the new concept of *perceived quality* for geospatial information. This solution also finds a new way to use the *Five Star Method* in a 3D geospatial environment which differs from the way in which the method is used in popular websites. The use of an extrinsic visualization method also validates the conclusions of MacEachren *et al.* (2005) where they assert that extrinsic visualizations are better as a method for communicating uncertainty for information from specific locations such as the 3D buildings used in the prototype, as the results of the user survey indicate that the users found this method easy to understand. The visualization method used in this research was tested using 3D buildings in Google Earth, but could be used for any other type of 3D object. Other future works could test the method using a different VG platform, and for different 3D objects.

5.3 Bibliography

Beard, K., Mackaness, W., 1993. Visual Access to Data Quality in Geographic Information Systems. *Cartographica*, 30:2&3, pp. 37-47

Bédard, Y., 1986. A Study of Data Using a Communication based Conceptual Framework of Land Information Systems. *Le Géomètre Canadien*. 40:4, pp. 449-460

Bédard, Y., 1988. Uncertainties in Land Information Systems Databases. *AUTO-CARTO* 8, pp. 175-184

Bédard, Y., Rivest, S. & Proulx, M. J., 2007. Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. Data Warehouses and OLAP: Concepts, Architectures and Solutions. In: Wrembel, R., Koncilia, C. (eds.), *Data Warehouses and OLAP: Concepts, Architectures and Solutions*. IGI Global. pp. 298-319

Bertin, J., 1981. *Graphics and Graphic Information-Processing*. Walter de Gruyter, New York

Budhathoki, N., 2010. Participants' motivations to Contribute Geographic Information in an Online Community. *PhD Thesis*. University of Illinois at Urbana-Champaign

Butler, D., 2006. The web-wide world. *Nature*, 439:7078, pp. 776-778

Buttenfield, B. P., 1993. Representing Data Quality. *Cartographica*, 30:2&3, pp. 1-7

Chrisman, N. R., 1983. The role of quality information in the long-term functioning of a geographic information system. In: Douglas D H (eds.), *Proceedings of AUTO-CARTO 6*. Ottawa, ON, Canadian Institute of Surveying and Canadian Cartographic Association: 303–21

Coleman, D. J., 2010. The Potential and Early Limitations of Volunteered Geographic Information. *Geomatica,* 64:2, pp. 209-219

Coleman, D. J., Georgiadou, Y., Labonte, J., 2009. Volunteered Geographic Information: the nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research*. 4, pp. 332-358

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: A position paper from the Vespucci Initiative for the Advancement of Geographic

Information Science. Editorial. *International Journal of Spatial Data Infrastructures Research*, 3, pp. 146-167

Davidson, E. Vaast, E., 2009. Tech Talk: An Investigation of Blogging in Technology Innovation Discourse. Professional Communication, *IEEE Transactions*. 52:1, pp. 40 - 60

Davis, P., Keller, C., 1997. Modeling and Visualizing Multiple Spatial Uncertainties. *Computers & Geosciences*. 23:4, pp. 397-408

Devillers, R., Bédard, Y., Jeansoulin, R., 2005. Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS. *Photogrammetric Engineering & Remote Sensing*. 71:2, pp. 205–215

Devillers, R., Jeansoulin, R., 2006. Spatial Data Quality: Concepts. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 31-42

Devillers, R., Beard, K., 2006. Communication and Use of Spatial Data Quality Information in GIS. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 237-254

Dokoupil, T., 2008. Revenge of the Experts. *Newsweek web exclusive*, March 6. Last accessed, June 2011.<http://www.newsweek.com/id/119091>

Drecki, I., 2002. Visualization of Uncertainty in Geographical Data. In: Shi, W., Fisher, P. F., Goodchild, M. F., (Eds.), *Spatial Data Quality*. Taylor & Francis, New York: NY, pp. 140-160

Drecki, I., 2007. Geographical Information Uncertainty: The Concept and Representational Challenges. *Proceedings from the 23rd International Cartographic Conference*. Moscow, Russia, 13 p

Editorial. 2006. Think global. *Nature*, 439:7078, pp. 763

Ehlschlaeger, C. A., Shortridge, A. M., Goodchild, M. F., 1997. Visualizing Spatial Data Uncertainty using Animation. *Computers & Geosciences*. 23:4, pp. 387-395

FGDC (Federal Geographic Data Committee), 1991. *Spatial Data Transfer Standard*. Washington, DC.: Department of the Interior

Fisher, P., 1999. Models of Uncertainty in Spatial Data. In: Longley, P., Goodchild, M. F., Maguire, D. J., Rhind, D. W., (Eds.), *Geographical Information Systems*. John Wiley, New York: NY, pp. 191-205

Gervais, M., Bédard, Y., Jeansoulin, R., Cervelle, B., 2007. Qualité des données géographiques. Obligations juridiques potentielles et modèle du producteur raisonnable. *Revue Internationale de Géomatique*, 17:1, pp. 33-62

Girres, J., Touya, G., 2010. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14:4, pp. 435-459

Goodchild, M. F., 2008. Spatial Accuracy 2.0. In: J.-X. Zhang and M. F. Goodchild (Eds.). Spatial Uncertainty. *Proceeding of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*. Volume 1. Liverpool: World Academic Union, pp. 1–7

Grossner, K. E., Goodchild, M. F., Clarke, K. C., 2008. Defining a Digital Earth System. *Transactions in GIS*, 12:1, pp. 145-160

Häberling, C., 2008. Proposed Cartographic Design Principles for 3D Maps: A Contribution to an Extended Cartographic Theory. *Cartographica: The International Journal for Geographic Information and Geovisualization* 43, pp. 175-188

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*. 37, pp. 682-703

Haklay, M., Singleton, A., Parker, C., 2008. Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*. 2:6, pp. 2011-2039

Hoeber, O., Wilson, G., Harding, S., Enguehard, R., Devillers, R., 2010. Visually representing geo-temporal differences. *IEEE Conference on Visual Analytics Science and Technology*, Salt Lake City, USA, 2010

Hunter, G. J., Goodchild, M. F., 1996. Communicating Uncertainty in Spatial Databases. *Transactions in GIS*, 1:1, pp. 13-24

Jones, M., 2007. Google's Geospatial Organizing Principle. *IEEE Computer Graphics and Applications*. 27: 4, pp. 8-13

Kim, S., Pantel, P., Chklovski, T., Pennacchiotti, M., 2006. Automatically Assessing Review Helpfulness. In *EMNLP 2006: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Sydney, pp. 423–430.

Kresse, W., Fadaie, K., 2004. *ISO Standards for Geographic Information*. Springer, Germany

Leyk, S., Boesch, R., Weibel, R., 2005. A conceptual framework for uncertainty investigation in map-based land cover change modeling. *Transactions in GIS*, 9:3, pp. 291-322

Lowell, K., 2004. Why aren't we making better use of uncertainty information in decision-making? *Proceedings of the Joint Meeting of Accuracy 2004 and TIES 2004*. 7 p

MacEachren, A., 1992. Visualizing uncertain information, *Cartographic Perspectives*, 13, pp. 10-19

MacEachren, A., Polsky, C., Haug, D., Brown, D., Boscoe, F., Beedasy, J., Pickle, L., Marrara, M., 1997. Visualizing Spatial Relationships Among Health, Environmental, and Demographic Statistics: Interface Design Issues, *Proceedings of the 18th International Cartographic Conference*. Stockholm, Sweden, 9 p

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, S., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32:3, pp. 139-160

Madry, S., Henley, A., Google Earth for Scientific Regional Analysis. 2007. University of North Carolina. July 7, 2009. <http://www.lib.unc.edu/reference/gis/ge>

Mangroves of Mexico as Google Earth Outreach example for GE6 in cooperation with CONABIO. 2010. In: *Landscape Visualization*. Last accessed, June 2011 <http://www.lviz.org/?p=349>

Nkhwanana, N. J., 2009. Assessing the Credibility and Trust of Volunteers of VGI. *White paper prepared for GEOIDE project IV- 41*. 23 p

Pang, A., 2001. Visualizing Uncertainty in Geo-spatial Data. In: *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*. pp. 1-14

Peter, M., 2009. Presentation and Evaluation of Inconsistencies in Multiply Represented 3D Building Models. In: Kurt Rothermel, Dieter Fritsch, Wolfgang Blochinger, Frank Dürr (eds.), *Quality of Context, First International Workshop, QuaCon 2009, Stuttgart, Germany, June 25-26, 2009*. Springer, pp. 156-163

Qu, Z., Zhang, H., Li, H., 2008. Determinants of online merchant rating: Content analysis of consumer comments about Yahoo merchants. *Decision Support Systems* 46, pp. 440-449

Rakshit, R., Ogneva-Himmelberger, Y., 2008. Application of Virtual Globes in Education. *Geography Compass*. 2:6, pp. 1995-2010

Riazanoff, S., Santer, R., 2006. Quality of Raster Data In Spatial Data. In: Devillers, R., Jeansoulin, R. (eds.), *Fundamentals of Spatial Data Quality*. Great Britain: ISTE Ltd. pp. 43-60

Roth, R. E., 2009. The Impact of User Expertise on Geographic Risk Assessment Under Uncertain Conditions. Cartography and Geographic Information Science. 36:1, pp. 29-43

Rouse, L. J., Bergeron, S. J., Harris, T. M., 2007. Participating in the Geospatial Web: Collaborative Mapping, Social Networks and Participatory GIS. In: Scharl, A., Tochtermann, K., (Eds.). *The Geospatial Web: How Geobrowsers, Social software and Web 2.0 are Shaping the Network Society*. Springer-Verlag. London, pp. 153-159

Schlaisich, I., Mountrakis, G., Agouris, P., 2004. Visualization of Image Quality in Distributed Spatial Databases. In: M. O. Altan (Editor), *International Society for Photogrammetry and Remote Sensing XX^th Congress*. Istanbul, Turkey.

Schroth, O, Pond, E., Muir-Owen, S., Campbell, C., Sheppard, S., 2009. Tools for the understanding of spatio-temporal climate scenarios in local planning: Kimberley (BC) case study. *SNSF Report PBEZP1-122976*. 33 p

Sheppard, S. R. G., Cizek, P., 2009. The ethics of Google Earth: Crossing thresholds from spatial data to landscape visualisation. *Journal of Environmental Management*. 90, pp. 2102-2117

Slocum, T., McMaster, R., Kessler, F., Howard, H., 2005. *Thematic Cartography and Geographic Visualization*. Pearson Prentice Hall, NJ, 518 p

Talking to DC. In Adam Bosworth's Weblog. Last accessed, June 2011. <http://adambosworth.net/2009/10/29/talking-to-dc/>

Taylor, F., Caquard, S., 2006. Cybercartography: Maps and Mapping in the information Era. *Cartographica*. 41:1, pp 1-5

Thoms, B., Garrett, N., Ryan, T., 2010. The Design and Evaluation of a Peer Ratings System for Online Learning Communities. In *HICSS 43: Proceedings of the 43rd Hawaii International Conference on System Sciences*, Koloa, Kauai, Hawaii, USA, IEEE, 10 p.

Tiede, D., Lang, S., 2007. Analytical 3D Views and Virtual Globes – Putting Analytical Results into Spatial Context. *Paper from: ISPRS Joint Workshop "Visualization and Exploration of Geospatial Data"*. 6 p

Tuttle, B. T., Anderson, S., Huff, R., 2008. Virtual Globes: An Overview of Their History, Uses, and Future Challenges. *Geography Compass*. 2:5, pp. 1478-1505

Uncertainty. 2009. In: *Merriam-Webster Online Dictionary*. Last accessed, June 2011 <http://www.merriam-webster.com/dictionary/uncertainty>

Warner, M., 2008. *Applied statistics: from bivariate through multivariate techniques*. SAGE Publications, Inc, CA

Wittenbrink, C. M., Pang, A. T., Lodha, S. K., 1996. Glyphs for visualizing uncertainty in vector fields. In the *IEEE Transactions on Visualization and Computer Graphics*. 2:3,pp. 266-279.

Zandbergen, P. A., 2009. Geocoding Quality and Implications for Spatial Analysis. *Geography Compass*. 3, pp. 1-34

Zuk, T. D., 2008. Visualizing Uncertainty. *PhD thesis*, Department of Computer Science, Calgary; Canada, 336 p

Appendix A: User Survey on geospatial data quality visualization

Privacy Statement: The personal information that you provide (your first and last names and your email address) will remain confidential with the distributors of this survey (Krista Jones, Rodolphe Devillers, Yvan Bédard and Olaf Schroth). The other answers that you provide for this survey will be compiled with those of others and published as part of an MSc thesis.

Please print your name and sign below if you agree to this privacy statement:

Name: ____________________

Date: ______________

Signature: _________________

Page 1 of 4

Survey on geospatial data quality visualization

Krista Jones: MSc candidate, Department of Geography,
Memorial University of Newfoundland

Date: ____________

Personal Information:

First Name: ______________ Last Name: ______________ Email: ______________

1. Are you familiar with what a virtual globe is (e.g. Google Earth)? Y☐ N☐

If you answer No to question 1 please return the survey uncompleted.

2. How frequently do you use virtual globes? Never ☐ Daily☐ Weekly ☐ Monthly☐ Yearly☐

3. Are you familiar with any GIS or mapping software (e.g. ArcGIS)? Y☐ N☐

4. Do you have any formal training or education in GIS? Y☐ N☐

Context

A number of Web mapping applications such as Google Maps and OpenStreetMap allow users to add and modify geospatial data. In a similar way, Google allows users to create 3D models in Google Sketchup and then make them available for anyone for download and use from their 3D Warehouse. These approaches are different from traditional geospatial data production methods as they are not based on explicit production standards, but can allow many users to contribute to the production of a single geospatial dataset. This creates challenges when you want to use such data to make a decision, as the quality of the geospatial dataset can be highly heterogeneous or even unknown. For instance, a user of a virtual globe could be immersed in an urban 3D environment, but have little indication of how reliable the different 3D models visualised are.

Five star method

To communicate the quality of 3D objects being made and used by all types of people in virtual globes, methods adapted from a number of other domains are being used in this project. These other domains have successfully implemented methods of communicating the reliability or quality of data/products to users. This includes popular Web sites or systems such as Amazon, eBay and ITunes which allow users to assess the quality of products using different rating and feedback systems. Most of these systems are based on a five-star rating approach and often allow users to add more detailed comments in relation to their own perception of the quality of these data/products.

The five star methods you will see in Part 2 of this survey are representative of an average rating that has been assigned to each of the 3D objects (buildings).

Part 1

a) View the four 3D buildings, one at a time, found under the heading 'Part 1' in Google Earth.

b) Consider, in your opinion, what the quality of these 3D buildings is.
Rank each of the four buildings from best to worst quality. Indicate the building (e.g. Building A) you consider to have the best quality above the 1, the building above the 2 as the the next best, and so on until you rank them all.

________	________	________	________
1	2	3	4

Part 2

a) View each of the four methods of visualizing the five star ranking method found under the heading 'Part 2' in Google Earth. View each of the four methods for Prague and San Francisco.
   - '1_Five point method', '2_Number of stars', '3_Layered stars', '4_Number within star'

b) View them in the order which was given to you (e.g. 4,2,1,3)

c) Consider, in your opinion, which of these methods is easiest to understand:

Five point method, Number of stars, Layered stars, Number within star

Rank each of the four methods from easiest to most difficult to understand. Indicate the five star method (e.g. Five point method) you consider easiest to understand above the 1, the five star method above the 2 as the next easiest, and so on until you rank them all.

________	________	________	________
1	2	3	4

d) Provide two reasons which you used in the selection of the best method (1st).
1) ____
____
2) ____
____

e) Provide two reasons you used in the selection of the worst method (4^{th}).

1) ____

____

2) ____

____

Part 3

a) View the four 3D buildings, one at a time, found under the heading 'Part 3' in Google Earth.

b) Consider, in your opinion, what the quality of these 3D buildings is.
Rank each of the four buildings from best to worst quality. Indicate the building (e.g. Building A) you consider to have the best quality above the 1, the building above the 2 to as the next best, and so on until you rank them all.

________	________	________	________
1	2	3	4

c) Did viewing the quality information about the 3D objects (buildings) affect your rankings of them? If so, explain why viewing the quality information changed your ranking of the buildings.

____

____

____

____