LNG&C

Manfred Buchroithner
Nikolas Prechtel
Dirk Burghardt *Editors*

Cartography from Pole to Pole

Selected Contributions to the
XXVIth International Conference
of the ICA, Dresden 2013

ICA
ACI

Springer

Lecture Notes in Geoinformation and Cartography

Publications of the International Cartographic Association (ICA)

Manfred Buchroithner · Nikolas Prechtel
Dirk Burghardt
Editors

Cartography from Pole to Pole

Selected Contributions to the XXVIth
International Conference of the ICA,
Dresden 2013

Springer

Editors
Manfred Buchroithner
Nikolas Prechtel
Dirk Burghardt
Institute for Cartography
TU Dresden
Dresden
Germany

Preface

The theme of the 26th International Cartographic Conference, "Cartography from Pole to Pole", illustrates the global importance of cartography in the analysis and presentation of spatial- and time-related issues. Many hundreds of years ago cartography may have had an impact on selected groups of a society only, it has, however, gradually and undisputedly evolved into something indispensable for the majority of mankind a long time ago. We might even speak of something indispensable on a nearly every-day basis.

Through the monitoring of an astonishing range of phenomena on earth from ground, from air and from space is obviously needed to react on natural dynamics and anthropogenous changes of environment to assess the impact of new technologies and economic activities on a globalised planet, geo-data are produced and recorded in an incredible volume. But spatial patterns cannot be understood as lists and tables of numeric data. It may be a commonplace for cartographers, however, only the knowledge, techniques and skills, which cartography has developed over its long history, will master the important step from data to understandable information. Cartographic products are ubiquitous nowadays, on paper, but even more so in the Internet. Cartography is the discipline which is concerned that these presentations will be more than colourful graphics. Contributions from more than 70 countries to the Dresden ICC conference clearly document that there exists a wide field of research, a strong demand for education, for practical realisation and for a dialogue with the users of cartographic products. The cartographic community will certainly do their best to accept this challenge and keep reporting on the progress.

By now it is the second time after the 25th International Cartographic Conference (ICC) 2011 in Paris that on the occasion of this biannual world congress for cartography a volume of the renowned Springer Lecture Notes on Geoinformation and Cartography (LNGC) is published. The Local Organising Committee (LOC) of the 26th ICC in Dresden, and in particular the editors of both the Digital ICC Proceedings and the Springer LNGC Volumes are proud, that out of approximately 820 entries submitted for this event the authors of some 170, who had right from the beginning written full-length papers, were obviously interested in having their typescripts published there.

Finally, a total of 34 have been selected for this book, based on a double-blind reviewing process organised by the LOC and coordinated by the ICA Commission Chairs. Around 180 international experts in certain thematic areas took up the major load of evaluating the abstract and full paper submissions. Those who have been concerned with the papers published in this book are listed right after this preface.

The articles were finally grouped into the 9 parts starting with *Cartographic Applications,* followed by *Cartographic Tools, Generalisation and Update Propagation*, *Higher Dimensional Visualisation and Augmented Reality*, *Planetary Mapping Issues, Cartography and Environmental Modelling*, *User Generated Content and Spatial Data Infrastructure, Use and Usability*, concluding with *Cartography and GIS in Education*. Since the book series of the Springer LNGC volumes have been experiencing a high international reputation, we can only hope that the papers published therein will be read by a multitude of not only cartographers.

It is the editors' pleasure to express their gratitude to Juliane Hermann, Daniel Horn, Peter Klengler, Steffi Sharma and Peggy Thiemt, all Dresden, for their support in the time-critical production of this volume. Without the experience of Agata Oelschläger from Springer Heidelberg, however, we would certainly not have succeeded to finalise this volume in due time. Many thanks go also to her.

Dresden, Summer 2013   Manfred Buchroithner
                                                                  Nikolas Prechtel
                                                                   Dirk Burghardt

International Scientific Committee

Suchith Anand
Gennady Andrienko
Masatoshi Arikawa
Thierry Badard
Sandrine Balley
Temenoujka Bandrova
Dirk Burghardt
Rex G. Cammack
Sébastien Caquard
William Cartwright
Xiaoyong Chen
Steve Chilton
Peter Collier
Antony Cooper
Alejandra Coll Escanilla
Philippe De Mayer
Imre Josef Demhardt
Qingyun Du
Cecile Duchene
Jason Dykes
Corné van Elzakker
Sara Fabrikant
David Fairbairn
Kenneth Field
David Forrest
Amy Griffin
Henrik Hargitai
Anja Hopfstock
Florian Hruby

Lorenz Hurni
Bernhard Jenny
Bin Jiang
Markus Jobst
Peter Jordan
Zhao Junqiao
Alexander Kent
Pyry Kettunen
Ralf Klammer
Milan Konecny
Alexandra Koussoulakou
Menno-Jan Kraak
Horst Kremers
Diah K. Kresnawati
Karel Kriz
Miljenko Lapaine
Rongxing Li
Elri Liebenberg
Yuefeng Liu
Christophe Lienert
Evangelos Livieratos
Lucia Lovison-Golob
William Mackaness
Carme Montaner
José Jesús Reyes Nuñez
Kristien Ooms
Karel Pavelka
Alastair Pearson
Chris Perkins

Mike Peterson
Maria Pla
Barbara Piatti
Nikolas Prechtel
Alexander Pucher
Nicolas Regnauld
José Jesús Reyes
Waldirene Ribeiro
Qingwen Qi
Manuela Schmidt
Jörn Seeman
Kira Shingareva
René Sieber
Stefan Steiniger
Vladimir Tikunov
Xiaohua Tong
Maria Tsakiri
Andrew Turner
Dražen Tutic
Lynn Usery
Necla Ulugtekin
Marita Wahlisch
Mark Ware
Robert Weibel
Lixin Wu
Xiaojun Yang
Xiaobai Yao
Lázló Zentai

Contents

Part V   Planetary Mapping Issues

Part VI   Cartography and Environmental Modelling

Part IX Cartography and GIS in Education

Contributors

Felicia O. Akinyemi Faculty of Architecture and Environmental Design, Kigali Institute of Science and Technology, Kigali, Rwanda, e-mail: felicia.akinyemi@ gmail.com

Mirela Altic Institute of Social Sciences, Centre for Urban and Local History, Zagreb, Croatia, e-mail: mirela.altic@zg.t-com.hr

Gennady Andrienko Knowledge Discovery, Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme, Sankt Augustin, Germany, e-mail: gennady.andrienko@iais.fraunhofer.de

Angelica Carvalho di Maio Universidade Federal de Viçosa, Viçosa, Brazil, e-mail: dimaio@vm.uff.br

Adriana Castreghini de Freitas Pereira Geosciences, Londrina State University, Londrina, Brazil, e-mail: adrianacfp@uel.br

Arzu Coltekin Department of Geography, University of Zurich, Zurich, Switzerland, e-mail: arzu@geo.uzh.ch

Antony K. Cooper Built Environment Unit, Council for Scientific and Industrial Research, Pretoria, South Africa, e-mail: acooper@csir.co.za

Tobias Dahinden Leibniz Universität Hannover, Institut für Kartographie und Geoinformatik, Hannover, Germany, e-mail: tobias.dahinden@ikg.uni-hannover.de

Paule-Annick Davoine Laboratoire d'Informatique de Grenoble, Grenoble, France, e-mail: Paule-Annick.Davoine@imag.fr

Daniel Eggert Leibniz Universität Hannover, Institut für Kartographie und Geoinformatik, Hannover, Germany, e-mail: eggert@ikg.uni-hannover.de

Sanet Eksteen Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa, e-mail: sanet.eksteen@up.ac.za

David Fairbairn School of Civil Engineering and Geosciences, Newcastle University, Newcastle, Great Britain, e-mail: dave.fairbairn@ncl.ac.uk

Hongchao Fan Chair of GIScience, University of Heidelberg, Heidelberg, Germany, e-mail: hongchao.fan@geog.uni-heidelberg.de

Stephan van Gasselt Planetary Sciences and Remote Sensing, Freie Universität Berlin, Berlin, Germany, e-mail: vanGasselt@gmail.com

Michael E. Hodgson Department of Geography, University of South Carolina, Columbia, USA, e-mail: hodgsonm@sc.edu

Kusay Jaara Cogit Laboratoiry, Institut Géographique National, Paris, France, e-mail: kusay.jaara@ign.fr

Zhao Junqiao Department of GIS Technology, OTB Research Institute, Delft University of Technology, Delft, The Netherlands, e-mail: johnzjq@gmail.com

Christine Lehmann Computer Graphics Systems, Hasso-Plattner-Institute, University of Potsdam, Potsdam, Germany, e-mail: christine.lehmann@hpi.uni-potsdam.de

Andrea Nass Department of Planetary Geology, Institute for Planetary Research, German Aerospace Center, Berlin, Germany, e-mail: andrea.nass@dlr.de

Milad Niroumand Jadid Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran, e-mail: milad.niroumand@yahoo.com

Maxim Nyrtsov Department of Geography of Geodesy and Cartography, Moscow State University, Moscow, Russia, e-mail: nyrtsovmaxim@gmail.com

Sebastian Pannasch Department of Psychology, Technische Universität Dresden, Dresden, Germany, e-mail: pannasch@psychologie.tu-dresden.de

Stefan Peters Technical University Munich, Munich, Germany, e-mail: stefan.peters@bv.tum.de

Marina Rajaković Faculty of Geodesy, University of Zagreb, Zagrab, Croatia, e-mail: mrajakovic@geof.hr

Fu Ren School of Resource and Environmental Science, Wuhan University, Wuhan, China, e-mail: renfu@whu.edu.cn

Na Ren Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing, China, e-mail: renna1026@163.com

Anne Ruas Institut français des sciences et technologies des transports, de l'aménagement et des réseaux, Paris, France, e-mail: anne.ruas@ifsttar.fr

Kisco M. Sinvula Department of Geography, Centre for Geoinformation Science, University of Pretoria, Pretoria, South Africa, e-mail: ksinvula@gmail.com

Dieu Tien Bui Faculty of Surveying and Mapping, Hanoi University of Mining and Geology, Hanoi, Vietnam, e-mail: Bui-Tien.Dieu@umb.no

Laurene Vaughan School of Media and Communication, Royal Melbourne Institute of Technology, Melbourne, Australia, e-mail: laurene.vaughan@rmit.edu.au

Tao Wang Future Cities Laboratory, Singapore-ETH Centre, Singapore, Singapore, e-mail: tawang@ethz.ch

László Zentai Department of Cartography and Geoinformatics, Eötvös University, Budapest, Hungary, e-mail: lzentai@caesar.elte.hu

Xiang Zhang School of Resource and Environmental Science, Cartography and Geoinformatics, Wuhan University, Wuhan, China, e-mail: xiang.zhang@whu.edu.cn

Part I
Cartographic Applications

GéoPeuple: The Creation and the Analysis of Topographic and Demographic Data Over 200 Years

Anne Ruas, Christine Plumejeaud, Lucie Nahassia, Eric Grosso, Ana-Maria Olteanu, Benoit Costes, Marie-Christine Vouloir and Claude Motte

Abstract The aim of the GeoPeuple project is to analyze the raise of population from the late twenty-eight to the early twenty-first century according to the topographic elements that characterize each commune (administrative area conceptually close to municipality): the infrastructure, the equipment, the settlements as well as the natural component such as the relief. We wish to learn more about the history of each commune but also to identify stylized facts if any. In order to understand the evolution of the population at the commune level, a first web interface has been proposed. It allows a better understanding of the aggregation processes. Then we built topographic vector data bases from old maps which required the understanding of the map content as well as a long process of interactive digitalization. To start the analysis step, we developed indices that characterize each commune. At least the analysis is performed: it is based first on the classification of each commune over time. Current study is focusing on the analysis of transitions over time.

Keywords Evolution analysis · Historical data bases · Demography

A. Ruas (✉)
LISIS Laboratory, IFSTTAR, Paris, France
e-mail: anne.ruas@ifsttar.fr

C. Plumejeaud · L. Nahassia · E. Grosso · A.-M. Olteanu · B. Costes
COGIT Laboratory, National Institute of Geographic and Forest Information (IGN), Saint Goussaud, France

M.-C. Vouloir · C. Motte
LaDéHis, École des Hautes Études en Sciences Sociales EHESS, Paris, France

1 Introduction

During the two last centuries, two main demographic mechanisms occurred in France: the raise of population (that went from 28 millions in the late XVIII to 66 Millions today) on one side, and the rural exodus which emptied the French countryside, on the other side We know that these phenomena are not spatially homogeneous: if the quantity of population globally growth, it did not grow in the same way, with the same ratio, at the same speed, everywhere (see for example Berry 1964 and Guérin-Pace 1993). We also know that there is a rank-size stability of cities over time (the Zipf's law) but that urban population growth is not only due to the size of the city and that interactions between cities should be taken into account to better estimate and understand the development of cities (Pumain 1997). The aim of this research is to look for criteria of differentiation in order to see if, with available topographic information, we are able to identify relationships between the quantity of population and the characteristic of each commune (city or village): the equipment, the infrastructure, the urbanization and the characteristic of the landscape. More precisely the aim of GeoPeuple project is to analyze the quantitative evolution of the French population, commune by commune, for two centuries according to the topography of each commune: the road network, the urban settlement, specific buildings—such as churches, factories, Water Mill—the hydrographic network, the relief and the administrative function of the commune. Thus we try to identify and characterize the relationships between the evolution of the quantity of population somewhere and the characteristic of this location. A first level of analysis consists in studying whether the differentiation/classification of the communes at a time stayed the same over time or changed. If it changed, are there any characteristics that could partially explain the evolution? A second level of analysis consists in looking for recurrent evolutions: are there transition patterns? If yes, can we find a partial explanation related to these patterns?

Both approaches help understanding the history of each commune and detecting some *stylized facts* (notion introduced by Nicholas Kaldor in 1961 and widely used in economics). This research requires crossing demographic and topographic data, and above all the constitution of historical geodata bases from existing sources. If historical demographic data already existed at the commune level in France, they had to be structured, linked and a web service has been developed to make this data accessible (*see* Sect. 2). The topographic data bases had to be created from old topographic maps (*see* Sect. 3). Finally demographic and topographic data were merged to make statistics at the commune level (Sect. 4).

GeoPeuple is a French research project, partially funded by the ANR that associates four French research teams: the COGIT from the IGN specialized in GIS, the LaDéHis from EHESS specialized in historical demography, the MALIRE from LIP6 specialized in image analysis and LISIS from IFSTTAR specialized in numerical models. In this chapter, we only present a part of the project focusing on the description of demographic and topographic data bases over time and the relationships between those data. Other chapter deal with the georeferencing process of

Cassini maps (Costes et al. 2012) as well as the proposed process to automate the detection of symbols on Cassini maps (Guyomard et al. 2012). Information is available of the project web site (http://www.geopeuple.ign.fr).

2 Demographic Data at the Level of the Commune

To avoid ambiguity we remind the meaning of *commune* in this article. A commune is the smallest administrative division in France. It contains a town hall and is managed by a mayor and a municipal council. A commune would correspond to a municipality in the USA and to a Parish or Town in the UK, however as a Commune has much more functions, we preserve this term instead of municipality. The set of communes is a partition of France: every location belongs to one and only one commune. A commune can be a very small village inside a large rural area (such as *Presles* in Fig. 12) or a full urban area such as *Grenoble*. The number of population is very heterogeneous: Paris as more than 2.2 million inhabitants but 86 % of the communes has today less than 2000 inhabitants. In 1793 France was composed of around 40,000 communes and today it is still composed of 36,680 communes: if numerous communes regrouping have occurred in the past (and are still occurring), we see the number of communes is equivalent. However the stability in numbers does not mean a spatial stability: overtime the boundaries of the communes have been and are still reshaped (some communes are aggregated or shared; some fields are transferred from one commune to another).

In order to study the evolution of the demography at the commune level, we use two data bases that are matched to constitute a demographic data base:

1. CASSINI data base—built by EHESS from diverse information such as the law bulletins, the official journals and the census from 1793 to today—which contains the name of each commune with associate population quantity, its evolution and the hierarchy between administrative levels.
2. All communes digital footprint from 1793, created by the demographer Hervé Le Bras.

The resulting new demographic data base contains for each commune its geometry and a structured inspired by the results of the GeoNomenclature project (Motte and Pelissier 2003; Vouloir and Motte 2008) which defined for example different divisions of space that exist in a past, an historical and unique codification to identify each commune and a codification that describe all kinds of evolution. This new data base is used to make the analysis of relationships between the demography of a commune and its topography (*see* Sect. 4) but it has also been made available by means of new web services. The aim of this new web service is to be able to easily navigate in the demographic data over time, querying a name of a commune and focusing on specific time frame. In order to understand the regrouping processes, a graph of history is also computed on line and mapped (Figs. 1 and 2).
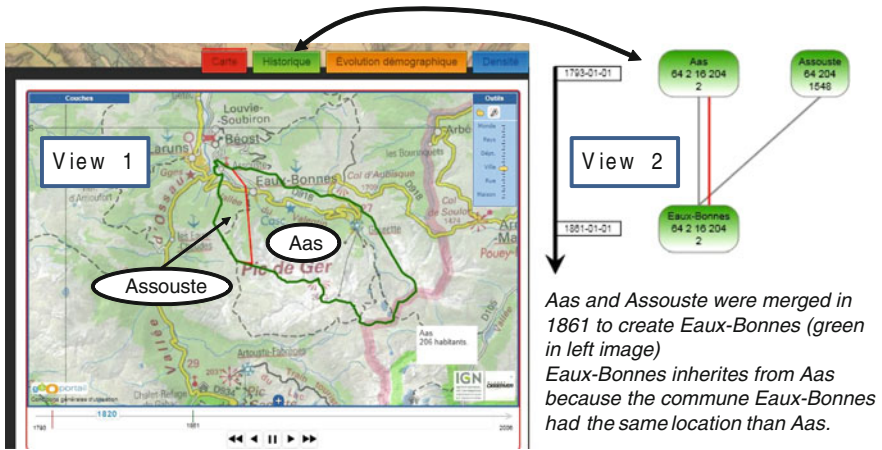
Fig. 1 Communes through the GeoPortail and the related historical graph
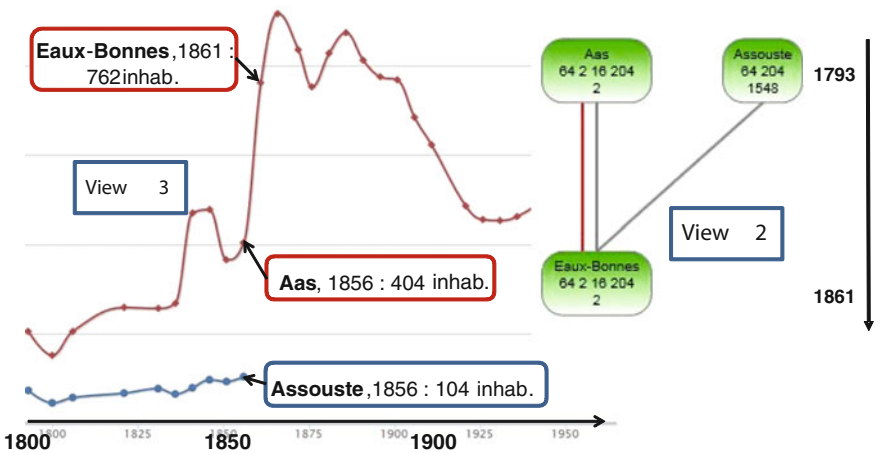
Fig. 2 Access to the evolution of the population for the 3 communes

Figures 1 and 2 are screenshots of the web interface that illustrates the story of three communes: *Aas* and *Assouste* were merged in 1861 to create the commune of *Eaux-Bonnes*. The communes boundaries are viewed on top of IGN GeoPortail (Fig. 1, *View 1, left*), the tag 'Historique' allows to view the historical graph of *Aas*, *Assouste* and *Eaux-Bonnes* (Fig. 1, *View 2 right* and Fig. 2, *View 2*) where as the tag 'Evolution démographique' shows the graph of population of these communes (Fig. 2, *View 3*).

The technical choices (SVG, Raphael, HighCharts, JQuery, API GeoPortail and OpenLayers) are presented and justified in (Grosso et al. 2012) and (Plumejeaud et al. 2012).

3 From Maps to Data Bases

A major difficulty of the project is to create data bases from old French maps. If we do have today topographic data bases from the IGN (the RGE), the other data bases had to be created. From the beginning of the project, it was decided to use traditional French topographic maps: Cassini maps for the end of twenty-eight century, Etat Major maps for the XIX, and IGN blue series for the mid twetieth century. These map series cover all France and have been made through well known processes. A first sight of these maps shows that the legend changed a lot, specifically between Cassini and Etat Major (*see* Fig. 3 *on an area that did not changed much*) whereas Etat Major and IGN blue Series look more similar. Four regions have been chosen for their topographic and demographic diversity (Saint Malo, Reims, Grenoble and Agen).

After having referencing Cassini maps (Costes et al. 2012), the team begun the interactive acquisition step to create topographic data bases. Actually it was not so simple as the Cassini and Etat Major data schema did not exist and should be created from an accurate analysis of maps and existing studies such as (Pelletier 1990). Figure 4 illustrates the difficulties of interpreting old maps. Symbols are small, hand drawing and sometimes difficult to distinguish or interpret. Moreover, many terms are no more used or their meaning is fuzzy for us. However a data schema has been set as well as a document to facilitate the interactive digitalization that can be viewed as specifications for data base acquisition.

Fig. 3 Representations of topographic data around Beine-Nauroy

Fig. 4 Cassini details: a chapel, two windmills and a church (from *left* to *right*)

Whenever data schemas were set, the digitalization could be performed (*see* Fig. 5 *for Cassini*).

Then a top-down data base alignment has been proposed in order to connect similar concepts in the different databases. The levels of detail being different, punctual objects from Cassini often correspond to surface objects in later data bases. For example *Carriage way* from Cassini are connected with *Road* from Etat Major; *Non Religious* and *Justice* from Cassini are connected to *Military* and *civil buildings* in Etat Major; *Railway* only exists from Etat Major. This matching
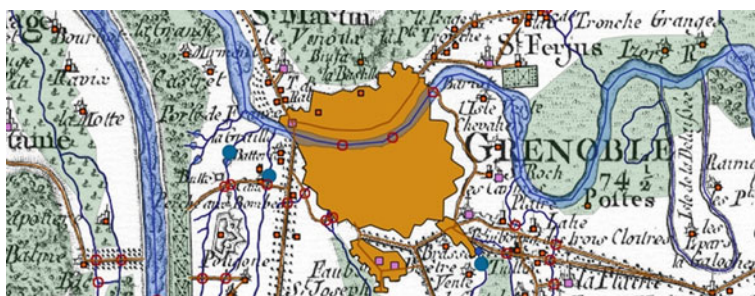
Fig. 5 Digitalization of Cassini maps

process is necessary to make statistics to describe the nature of each commune over time. Details of criteria to characterize each commune are given in Sect. 4.2.

A complementary study consists in matching similar objects one to another to learn about the *history of objects* at a very local level. For example, if many churches exist for a long time, how did wind or water mills evolved over time? This research is still ongoing; results will be published later on.

4 Relationships Between Demography and Topography at the Commune Level

4.1 Heterogeneous Urban Growth

The first very simple study consists in merely mapping urban areas at different period to see the urban growth.

Figure 6 illustrates an example in the suburb of Reims. Visually we can see that some communes did not growth so much (*Avenay*), others grew between Cassini and Etat Major (*Bisseul*), other between Etat Major and today (*Dizy*), while others seamed to growth more gradually (*Mareuil, Ay, Epernay*). The religious buildings from Cassini maps illustrate the structure of the distribution of the communes at
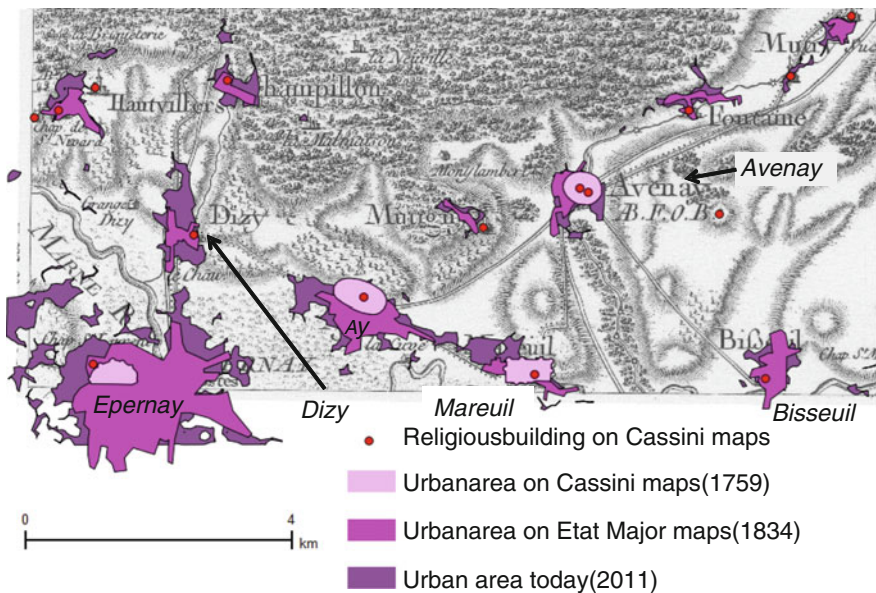
Fig. 6 Urban growth over time

least in the twenty-eight century. We can see that they coincide with the core of urbanization.

4.2 Criteria of Characterization

To characterize each commune a set of very classical indices have been chosen and implemented. We divide them into 5 categories:

1. *Population*. Number of population, density;
2. *Infrastructures*. Length of road, length of connecting road, length of river, length of channel, number of bridges, number of station;
3. *Settlements*. Urban area, number of scattered settlements;
4. *Equipment and activities*. Number of industrial buildings, number of religious buildings, number of military buildings, number of post-offices, administrative level of the commune;
5. *Landscape*. Average altitude and average slope.

The infrastructure gives contextual information as it allows having indictors on the connections between a commune and the others around.

When computing indicators we have to deal with two difficulties, one is related to the regrouping process, and the other is related to the different levels of detail of the information according to the data source (Fig. 7).

For the first point, we choose today commune boundaries as being the reference as today data are more accurate than the others. Thus the computation of quantities defined above is done inside today boundaries.

Road length 3,2 km in the late XVIII   8,7 km in the early XXI
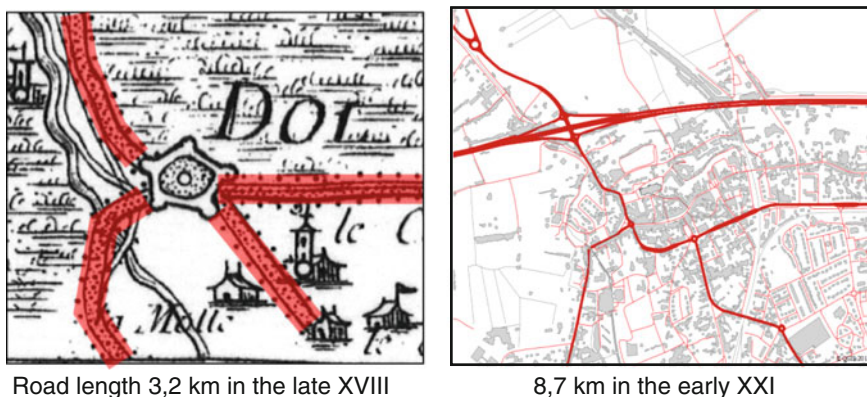
Fig. 7 Computing road length at Dol-de-Bretagne (Saint-Malo)—Cassini map and IGN BD TOPO source

For the second point, we had to study case by case what could be the relevant correspondence between concepts over time. For instance, concerning road networks, Fig. 7 shows the selection that has been made in order to compute roads length, with resulting road length computed on the same area.

4.3 Classification of Communes

The classification of communes is done with the criteria defined in Sect. 4.2. The classification is performed in two steps. The first consists in finding relevant set of criteria by means of a Principal Components Analysis (PCA) and the second is the classification itself, a Hierarchical Ascendant Classification (HAC) based on the selected criteria.

This methodology, illustrated in Fig. 8, inspired by work of (Piron et al. 2004) about socio-economic evolution of demographic groups living in Bogota district, is transposed to our case. A special care must be given to comparability of indicators for each times (t1 is Cassini and t2 is 2011), since they are not identical. Following the theory developed in (Lebart et al. 1995), distributions of various indicators are compared, in order to check whether their shape looks the same: it is the case. Our classification is using Euclidian distance, on a Pearson correlation matrix.

Two experiments have been performed:

- The first experiment consists in classifying separately the communes by dates and analyzing the resulting commune profiles. The point here is to verify how our methodology can discover the organization of space at various times, independently of any diachronic study.
- The second experiment consists in classifying the communes all together (with different dates) to illustrate the global evolution of communes.

Figure 9 shows results of the first experiment. Five classes have been found, and the majority of territorial units belong to the class in green on map, grouping units with the following characteristics: *widely dispersed areas of housing, located in low altitude, having little industry*. Their profile is opposed to the ones
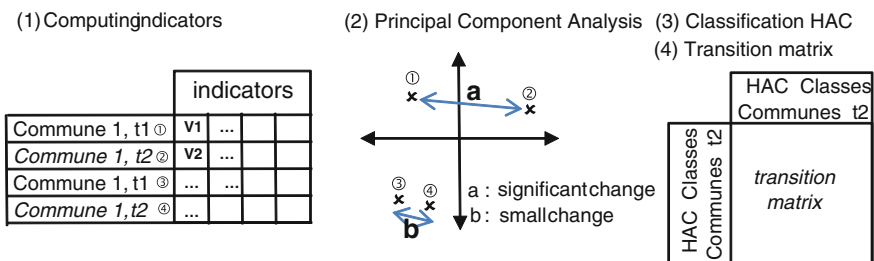
Fig. 8 Multivariate analysis methodology for building transition matrix
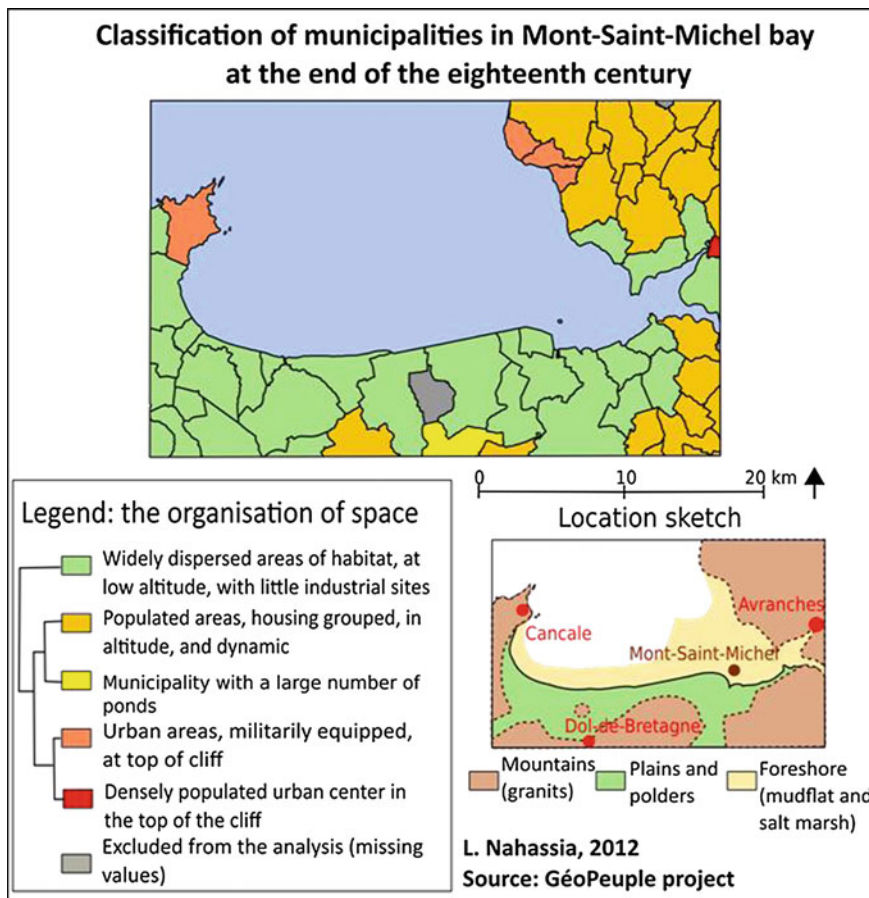
Fig. 9 Classification of communes around *Saint-Malo* during Cassini period

belonging to classes in orange or red, that are located in altitude (on cliff), and being more urbanised already at that time. This spatial organization reflects overall the geomorphologic profile of this region, which has not still been affected by all spatial planning projects of the twentieth century, aiming at canalising the water, and building dikes to protect housing from water floods.

In the following, results of the second experiment are displayed and commented. Figure 10 shows the extreme case with data from Cassini classified all together with 2011 data. All units under Cassini seem to belong to the same class (in dark green), corresponding to *dispersed areas of habitat, sparsely populated*, whereas units at twentieth century belong to various classes: urban ones, very populated, (in orange or red) or rural ones, (in dark or light green color).
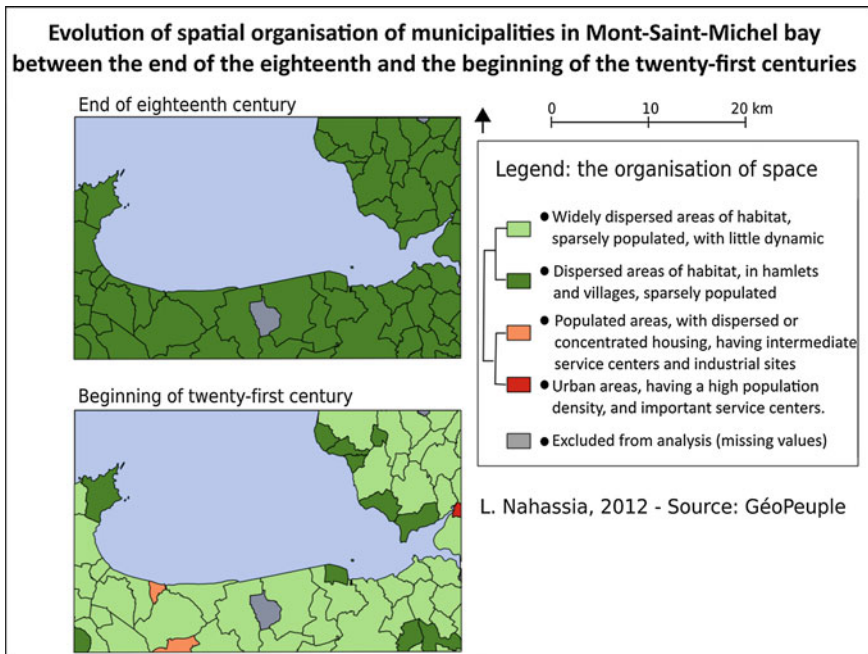
Fig. 10 Evolution of communes around *Saint-Malo*, from Cassini period to 2011

This methodology has been tested separately for each zone of our study, and it gives the same results for all of them. For instance, on *Grenoble*, Fig. 11 shows that, again, Cassini period seems to be quite homogeneous comparatively to 2011 year, while urban area is polarizing all space at 2011.

This means that a real densification process has occurred around urban areas, along with a rural exodus, in such proportions that most of past urban areas would only look like small villages for us today. Of course we know that communes where not identical in the past and if Cassini's indicators are by themselves sufficient to reveal differentiations at their time, they are completely crushed by BD TOPO's indicators, whose values are much more high/intense in term of roads, populations, or housing for instance.

It is then interesting to compare communes which change over time with the ones that stay in the same category. For instance, *Presles* (Fig. 12) remains a rural municipality all along the time, in the same class, whereas *Eybens* (Fig. 13) has changed and populated considerably since the eighteenth century. The main reason is that *Eybens* belongs to the suburb area of *Grenoble*, the Alpes's capital, attracting many services and industries, well served by roads and train, whereas *Presles* is located in altitude inside the *Vercors*, in a very high landscape with little accessibility by road, and only few farms to host tourism.

Evolution of spatial organisation of municipalities in Grenoble area between the end of the eighteenth and the beginning of the twenty-first centuries
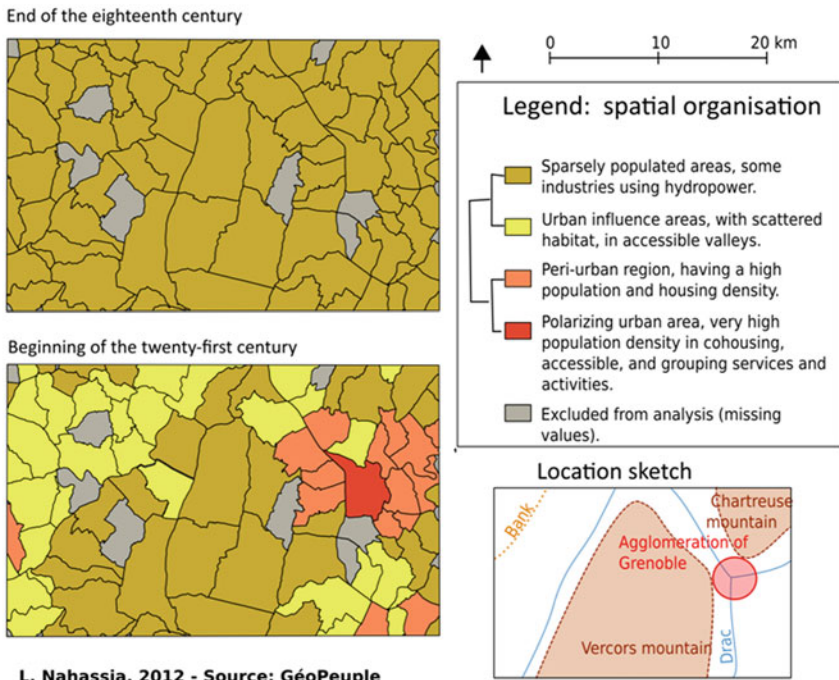
Fig. 11 Evolution of communes around *Grenoble*, from Cassini period to 2011

Fig. 12 *Presles* from Cassini to 2011
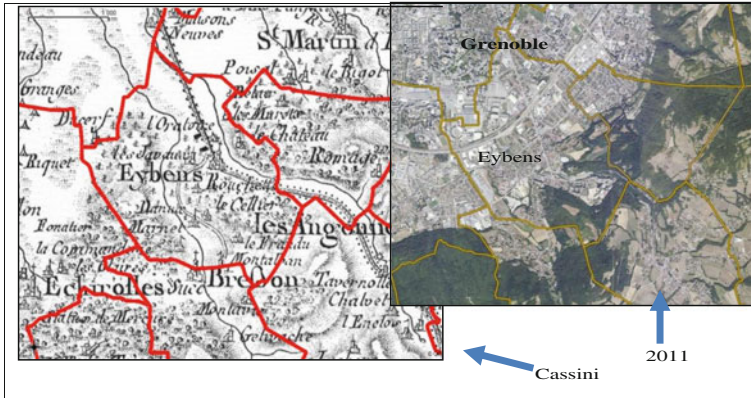
Fig. 13 *Eybens* from Cassini to 2011

4.4 Towards an Analysis of Transitions

The aim of this analysis is to look for transition patterns in order to answer to the following questions:

- Are there recurrent sequences of steps the communes are following? Which sequences?
- Which communes are meant to follow which sequences? Where are there?
- Are there relations between these sequences and the evolution of the population?

To make this study we introduce the following terms. We name:

- *Commune-type*. A classification of communes according to a typology based on some criteria defined in Sect. 4.2: 1/all criteria, 2/only the topography.
- *Commune-Transition*. A specific succession of commune-types for a specific commune.
- *Commune-Transition-pattern*. A generic succession of commune-types.
- *Population-Transition*. A specific sequence of population properties (quantity of population, density, augmentation rate) for each commune.
- *Population-Transition-pattern*. A generic succession of population properties.

Thus three kind of *transition patterns* can be seek: one set including all criteria (*topography and demography*, *see* classification 1 *in* Fig. 11), one set including only topography (classification 2) and one set including only demography (classification 3).

From these transition patterns two studies should be performed:

1. we can map the communes according to the classification 1 (all criteria) to study the spatial distribution of these transition patterns
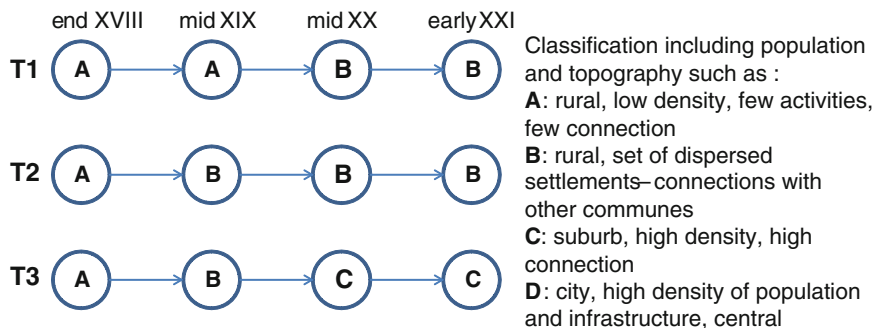
Fig. 14 Examples of 3 transition patterns T1, T2, T3 (classification 1)

2. We can study the correlations (or at least the relations) between the Communes-transition patterns (evolution of topography classification 2) and population-transition-patterns (classification 3).

These studies are particularly interesting as we may wonder as far the activities of a commune are linked to its infrastructure, equipments and settlements. Actually the existence of road does not mean the traffic is high, the existence of settlements does not mean that people are still living in the commune. This analysis is inspired from (Petitjean et al. 2012) developed during the GeOpenSim project (Ruas et al. 2011). This experiment illustrated in Fig. 14 has been performed in summer 2013.

5 Conclusion

The GeoPeuple project aims at analyzing the evolution of topography and demography all together at the commune level. The point is this to look for topographic elements, either anthropic (settlements, infrastructure, equipment) or natural (hydrography, relief) that could help understanding the differentiation of population growth. The demographic data at the level of the commune were structured and made available through a new and dedicated web services (Sect. 2). A difficulty of the project was to create data bases from old maps as data schema did not exist (Sect. 3). Then whenever the digital data set exist, we had to develop criteria and compute indices on each commune, at each time, taking into account the fact that the concepts and the level of detail are different from 1790 to today (Sect. 4.2). In order to analyze the evolution, two methods were proposed. The first is based on the analysis and comparison of commune classification (Sect. 4.3) (Nahassia and Plumejeaud 2012) whereas the second is based on the detection of classical sequences of states communes goes through that we call transition (Sect. 4.4).

The fact this kind of data are now available on a least 3 regions of France from 1790 to today opens the way to fundamental research, looking for a better understanding of evolution processes.

References

Berry B (1964) Cities as systems within systems of cities. Reg Sci 13(1):146–163

Costes B, Grosso E, Plumejeaud C (2012) Géoréférencement et appariement de données issues des cartes de Cassini—Intégration dans un référentiel topographique actuel. SAGEO 2012

Grosso E, Plumejeaud C, Parent B (2012) GeoPeuple project: using RESTful Web API to disseminate geohistorical database as open data. In: Proceedings of the 2nd open source geospatial research and education symposium (OGRS), Yverdon-les-Bains, 24–26 Oct

Guérin-Pace F (1993) Deux siècles de croissance urbaine. Anthropos, Paris

Guyomard J, Thome N, Cord M, Artières T (2012) Contextual detection of drawn symbols in old maps. In: International conference on image processing (ICIP)

Lebart L, Morineau A, Piron M (1995) Statistique exploratoire multidimensionnelle. Dunod, Paris

Motte C, Pélissier J-P (2003) Géonomenclature des lieux historiques habités. Archives de France, Paris

Nahassia L, Plumejeaud C (2012) Analyse spatio-temporelle de l'évolution des communes françaises depuis la révolution française à aujourd'hui. Manipulation et exploitation d'une base de données géo-historique. GeoPeuple Report number 4.1. http://geopeuple.ign.fr

Nicholas Kaldor (1961) 'Capital Accumulation and Economic Growth.' In: Lutz, Hague(eds) The Theory of Capital, London, pp 177–222

Pelletier M (1990) La Carte de Cassini. L'extraordinaire aventure de la carte de France. ENPC Press, Paris, p 263

Petitjean F, Inglada J, Gançarski P (2012) Satellite image time series analysis under time warping. In: IEEE transactions on geoscience and remote sensing

Piron M, Dureau F, Mullon C (2004) Utilisation de typologies multi-dates pour l'analyse des transformations socio-spatiales de Bogota (Colombie). Cybergeo Eur J Geogr, http://cybergeo.revues.org/3742

Plumejeaud C, Grosso E, Parent B (2012) Site web interactif pour l'exploration de l'histoire démographique des communes françaises de 1789 à aujourd'hui. GeoPeuple Report number 6.0. http://geopeuple.ign.fr

Pumain D (1997) Settlement size dynamics in: history. In: Holm E (ed) Modelling space and networks. Umea, GERUM, Proceedings from 7th colloquium of theoretical and quantitative geography, Stockholm, Sept 1991, pp 3–32

Ruas A, Perret J, Curie F, Mas A, Puissant A, Skupinski G, Badariotti D, Weber C, Gancarski P, Lachiche N, Lesbegueries J, Braud A (2011) Conception of a GIS-platform to simulate urban densification based on the analysis of topographic data. In: 25th international cartographic conférence. Lectures notes in geoinformation and cartography, vol 2. Springer, Heidelberg, pp 413–430

Vouloir M-C, Motte C (2008) Frontières administratives et identités communales. Le cas de la France, XVIIIe–XXe siècles. The historical review, Athènes, vol V

Implementation of Cartographic and Digital Techniques in Orienteering Maps

László Zentai

Abstract Orienteering maps are very special maps, because not only they are using the same specification all over the World, but also the users themselves create these maps regularly. This paper summarizes the implementation of the most important cartographic techniques of the last decades including the application of the information technology.

Keywords Orienteering maps · Topographic maps · Map drawing

1 Short History of Orienteering

Johann GutsMuths was a teacher and educator in Germany in the second part of the eighteenth century and in the beginning of the nineteenth century. He was especially known for his role in the development of physical education, so he is known to be the Great Grandfather of Gymnastics. In his famous book (Turnbuch für die Söhne des Vaterlandes, 1817, Frankfurt), he set great store by the open-air exercises including such activities as map reading and distance estimation.

The Swedish Military Academy already specified in 1840 that cadets should be familiar with the knowledge of distance and height measurements and they should be able to make map sketches on the terrain.

Orienteering as a sport started as a military navigation test/training at the end of the nineteenth century mostly in Norway and Sweden (the navigation as an open-air physical exercise became part of the military training curriculum around 1880). The word "orienteering" was used to mean crossing unknown territory with the aid of a map and compass for the first time in 1886. The first civil orienteering

L. Zentai (✉)
Department of Cartography and Geoinformatics,
Eötvös Loránd University, Budapest, Hungary
e-mail: lzentai@caesar.elte.hu

competition was held in *Norway* in 1897 (Zentai 2011). Some other countries (Denmark, Estonia, Finland, Hungary, Switzerland) organized their first orienteering events between the two world wars.

In the early 1930s, the sport received a technical boost with the invention of a new compass, more precise and faster to use.

Due to the slow development of the sports at that time, the first independent national orienteering federations were formed only around 1937–1938 in Norway and Sweden (Berglia et al. 1987).

1.1 Early Maps

The early period of orienteering maps was the age of using existing large scale state topographic maps. In the Scandinavian countries, these large scale maps were not classified in the beginning on the twentieth century. In other countries, however, there were no suitable maps available for public use. The running speed of competitors and the available large scale state topographic maps partly effected also the length of the courses.

The development of the sport was very slow after the first civil event; the main change was that Sweden took over the leading role of the sport from Norway. Concerning the maps of the orienteering events, the organizers had no other opportunity than using various official maps: state topographic maps, tourist maps. At this level of the development of the sport, there was no demand for special maps partly because the mapmaking was a very complicated and expensive process at that time and partly because the number of competitors was too small to cover the costs of producing such maps. The only change was the re-drawing of the existing topographic map in a larger scale, but without changing the content. However, sometimes all existing maps were so outdated for orienteering events that major revisions or updating would have been requested, but without experts and affordable methods this request could not be fulfilled.

The first orienteering map that was especially drawn and field-worked for orienteering was made in 1941 in Norway. The main reason of this special map was the inaccessibility of maps during the German occupation. During WW II, the sales of maps were stopped in most Scandinavian countries.

2 New Cartographic Techniques in Producing Orienteering Maps

The most challenging element of the large scale topographic maps at time was the representation of the relief. Although the experiences of the modern wars, especially WW I, proved that contour lines are the best method for this, but to

transform all existing hachured maps to a modern contour line representation was a very slow process. As the precision requirements of the state users increased, this required more time on field-working using the traditional methods (plane table, theodolite). The only solution to speed up the mapmaking process was to invent better and more efficient measuring methods.

2.1 Stereophotogrammetry

Topographic mapping using photogrammetry was introduced at the end of the nineteenth century. The first terrestrial cameras made stereo-photogrammetry possible in mountainous terrain before the airplane offered itself as a useful camera platform at the end of the nineteenth century. This terrestrial method made easier the mapping of mountainous areas (especially the representation of the relief), but this method was not suitable for the efficient mapping of larger areas.

In 1948, Norway made the first orienteering map where the contour lines were created from a special photogrammetric plot. On the very detailed Scandinavian terrains this method was used continuously, but due to the high costs of the stereophotogrammetry the spread of this method in orienteering maps was relatively slow in the beginnings. This method also allowed much more accurate relief representation on orienteering maps, which also let the event organizers use relief details as potential control point sites. To have more details, more features to represent on the orienteering map, it also caused to increase the scale in order to keep the legibility of the map. Using larger scale also meant that the production of orienteering maps started to be independent from the state topographic maps. In the Scandinavian countries, the largest scale of the state topographic maps was just increased to 1:25.000 at that time, but the scales of orienteering maps started to increase rapidly.

From the early fifties, major Scandinavian (especially Norwegian) clubs and individuals experimented with mapping. A large number of maps were made of varying standard. Many of the people involved were professional mapmakers. Some keen orienteers became familiar with stereophotogrammetry getting experience in stereo plotting from major mapping companies and higher education institutes. The first orienteering base maps prepared by private orienteering firms were produced in 1954–1955.

The revolution continued in the 1960s with the creation of the first orienteering map with a 5 m contour interval, where the contours were incredibly detailed. Orienteers had to accommodate their navigational skills to these special orienteering maps, but it was just the right time as the first European Orienteering Championships was organized in 1962. The maps contained so much information that a completely new orienteering technique was needed. The map was so detailed and accurate that the compass was not the most essential tool anymore; the competitors mostly just read the map.

2.2 Colour Offset Printing

Although we had suitable methods to make precise maps for orienteering, there was another crucial problem to be solved. Colours are essential parts of topographic maps, but the colour printing methods were expensive. Even the state topographic maps started to be printed in colour only from the beginning of the twentieth century, when colour offset printed methods were invented.

The very first colour offset orienteering map was produced in 1950 for an international event. For several years, colour maps were used only at the largest events, where the number of participants let the organizers finance this method. The scale of this first map was 1:20,000, with four colours; forests in light green, man-made features in black, water features in blue and contour lines in brown. The base map was Oslo city maps in the scale 1:5,000, with 5 m contours. The map was much more detailed than anything used in Scandinavia up to that time.

The colour printing also helped the fairness of the sport: since then, the better legibility helped that all the events became based on fine navigation with precise maps. Namely, the control points could be found by the ability to navigate, not by luck or searching.

2.3 Modern Map Drawing Tools and Techniques

One of the most special characteristics of orienteering maps is that they are made by orienteers and not by professional cartographers (although there are some cartographers who are keen orienteers). As mentioned in the previous papers, orienteering started to use the most modern base map creation methods (stereophotogrammetry) and printing techniques, which required not only professional experiences, but very expensive machines.

Other sections of the orienteering map production did not require expensive machines, but special experiences and tools were necessary. As orienteering maps started to become special maps, updating the base maps also meant that the maps had to be redrawn. Map drawing required special skills: dexterity, precision, material intensity, cleanness, but the perfect knowledge on orienteering was much more important. This caused several poorly drawn, ugly though relatively precise orienteering maps. After drawing some maps, a few orienteers showed affinity to this job and started to improve their drawing skills. The drawing quality of their maps became professional.

This drawing process was supported by certain tools as well: translucent plastic films, technical pens, drawing templates, stencils, dry transfer screens and letter transfers. These tools and materials were developed for technical drawings and cartographic production, but orienteers were open to adopt them for drawing orienteering maps. In a later stage, some special products were developed for the orienteering mapping: drawing templates, dry transfer screen and symbol sets.

Most of the professional cartographic firms used scribing for map production. Scribing was used to produce lines for cartographic map compilations around 1960–1990. The lines produced by this technique are sharp and clear. This technology required investments too (light table, translucent coating film, scribing tools: tripod, stylus). Scribing produced a result superior to drafting, but was more time consuming and required professional experience. Only few orienteering maps were made with this method, especially by small professional cartographic firms.

2.4 Standardization

The *International Orienteering Federation* (*IOF*) was founded in 1961 with ten countries. In 1964, the association of Nordic countries had formed a map committee and they asked the IOF to discuss the maps of international events and to form the Map Committee of the IOF. The first internationally accepted principles were as follows:

- The maps have to be so accurate and detailed that they give the possibility for the organizers to make a fair event.
- The main disadvantage of using outdated topographic maps without special orienteering fieldwork is the luck factor. If a competitor has found a path not shown on the map and has been able to use it during the event can reach the control point faster than the unlucky rivals who omitted that path.

In 1965, the IOF formed its Map Committee, but the progress was directed by the Scandinavian countries. All five members of the Map Committee were cartographers and orienteers (*Jan Martin Larsen*—Norway, *Osmo Niemelä*—Finland, *Christer Palm*—Sweden, *Torkil Laursen*—Denmark, *Ernst Spiess*—Switzerland).[1]

The most important and urgent work of the committee was the specification of World Championship maps:

- The orienteering maps have to be new.
- The map has to show every important detail of the terrain which can affect the route choice of the competitor.
- The most relevant characteristic of an orienteering map is the accuracy and legibility: small and unimportant details have to be omitted.
- The maps of international events have to use the same specification.

The first issue of the ISOM was released in 1969. This issue was still not a real specification but rather a "guideline", although it already contained quite concrete requirements (Spiess 1972) (Table 1).

---

[1] Ernst Spiess got the ICA Carl Mannerfelt Gold Medal in 2005; Christer Palm is an ICA Honorary Fellow (1997).

Table 1 Summary table of the international specification of orienteering maps

Year of publishing	Number of signs	Suggested scale
1969	52	1:25,000, 1:20,000
1975	100	1:20,000, 1:15,000
1982	98	1:15,000, 1:10,000
1990	105	1:15,000, 1:10,000
2000	104	1:15,000, 1:10,000

The main reason of the standardization was the globalization of the sport. It was initiated by not only the top level events (World Orienteering Championships), but since the 1970s more and more multi day events were organized, where more and more foreign participant took part.

It is also interesting to mention another aspect of globalization that is not directly related to mapping. Multi day events lead to linguistic problems. Control descriptions previously were given in written form, mostly in German at international events, because this was the official language of the IOF (it was changed to English in 1985, when the IOF had several non-European members, where German was hardly spoken). These descriptions had to be translated into several languages, often with strange results, as the translator had little knowledge of the orienteering terminology in various languages. In 1974, Swedish orienteers came up with a solution to this with the pictorial control descriptions, where all possible control sites were given a symbol resembling the IOF map symbol, but with the restriction that it should be reproduced in black and white (to make it easier to copy/print). The idea was adopted very quickly by the IOF, and nowadays these symbolic control descriptions are used even at small local events, where no foreign participants are expected.

2.5 Desktop Mapping (Computer Drawing)

The relatively small number of symbols in orienteering maps made it relatively easy to make formerly hand drawn orienteering maps by computer. The first vector based general graphic software for personal computers (Adobe Illustrator) and the first GIS software was released at the end of the 1980s, but professional companies, institutes (national mapping authorities) could use such software even earlier.

It is not easy to find the first orienteering map which was made by computer. Keen orienteers, employees of the *Norwegian Mapping Authority* made the first digital orienteering map around 1980, which was made by automated scribing in a drawing machine based on digitized field work.

The *Swedish Orienteering Federation* started a project in 1989, which laid the foundation for the future production of orienteering maps with digital methods. Based on this preparation, the federation started the centralized and targeted work to turn the Swedish orienteering mapping base into digital format in 1991.

Involved in this process were not only sports, but also authorities such as the *Swedish National Land Survey*, the *Communities Confederation* etc. It took about 5–7 years until the entire yearly production of orienteering maps was done with digital means as products of databases.

In *Finland*, in 1990 *Risto Laiho* made some A4 size maps with FINGIS (FINnish Geographic Information System) program, which was developed by the *National Board of Survey*. The application for orienteering maps was worked out by Risto Laiho. Although the development of FINGIS started in 1975, it was very difficult to use. FINGIS software was applied to terrain maps, geological maps, communal maps, forestry taxation, water resources management, cadastral and real estate surveys.

It was extremely difficult to draw all the symbols of the orienteering map specification in GIS software environment, especially because these systems were not developed for professional output.

In *Denmark*, *Flemming Nørgaard* managed to draw the first digital Danish orienteering map, which was published in April 1989 (Fig. 1). He also made the map of the first IOF event which was organized to use digital orienteering map (World Cup event, 1990: Gjern Bakker). These maps were made with Illustrator 88 on Apple-McIntosh personal computers.

Text is not an important feature of orienteering maps. Text elements are supplementary, but they do not really increase the usability of the maps for the competitors. This fact made it relatively profitable to create orienteering map drawing software. *Hans Steinegger*, a Swiss orienteer/software engineer released the first version of his *OCAD* software in 1988–1989. Nowadays, most of the orienteering maps all over the world are drawn with this software. At that time, the text handling features of the PC software were very limited due to the lack of standardisation. Only Windows 3.1 and later versions solved this problem at the beginning of the 1990s with using TrueType fonts.

OCAD became even more popular when scanners and colour inkjet printers became easily affordable and were substituted for the less comfortable input device, the digitizing tablet.

One of the most difficult tasks was to make the traditional orienteering mappers used to computers, to be familiar with the personal computers. Keen Apple personal computer users had no chance to use OCAD (at least not in the first part of the 1990s, when OCAD ran under DOS environment). On Mac computers, Adobe Illustrator was used producing the first digital orienteering maps on personal computers. One of the most difficult challenges for the beginner computer users was to be familiar with the output section of the orienteering mapping (desktop publishing, imagesetting, Postscript).

The user friendly environment of the OCAD software makes map drawing relatively easy even for beginners who are not familiar with any type of computer software. The development of colour printing technologies and the price reduction of good quality colour laser and ink jet printers made the colour printing technology more affordable.

Fig. 1 Risskov, Denmark. The first orienteering map made on Apple-McIntosh personal computer. Courtesy by Flemming Nørgaard

The number of offset printed maps is continuously decreasing and the number of non-offset printed maps is increasing. The main factor of this change is not simply financial, but the opportunity to very late updates before the event. The mapmaker is also able to control the complete printing process. The continuous development of colour printing technologies will continue and the quality of digital printed maps can approximate the quality of spot colour offset printed maps.

2.6 GPS

One of the most interesting parts of the orienteering mapping which requires orienteering experience is the *field-working*. Even if we have very good base maps, the whole area should be thoroughly checked and all base map information should

be transformed into the features of the orienteering map or partly dropped. Depending on the quality of the base map and the complexity of the terrain, the field-working should take about 30–50 h/km^2.

Global Positioning System devices are more commonly used during ground survey in general. To enable the data to be used easily, maps need to be "georeferenced". Orienteering maps are not regularly "georeferenced", which means that only very few of these maps were fitted to known projections and/or datum of the national or international mapping systems (datum, projection etc.). Theoretically, the orienteering maps used the same projection/datum as the original base maps, but as time went on and the old maps were updated, new areas were added with smaller and smaller distortions, and they were incorporated in the maps. The "unreferenced" orienteering maps were suitable for the events, because the inaccuracies were distributed on the whole area of the map, and these failures practically did not affect the navigation of the competitors, who use only the orienteering map and compass on the terrain. Absolute positional accuracy is of little significance compared to relative accuracy and to the proper representation of the terrain shape and features. Coordinates are not indicated in orienteering maps and GPS does not have a role in classic orienteering (according to the competition rules, external help during the events is prohibited for the competitors).

The "georeferencing" of existing maps is a time consuming process and requires some technical knowledge. This explains why orienteers need experts to help the "georeferencing". However, orienteering map projects normally do not allow financing such service.

The main advantages of using GPS:

- They are definitely more accurate than traditional surveying techniques (pace counting, bearing).
- Absolute positions are very helpful to improve mapping and can save the survey time.
- Easy to discover the base map errors and uncertainties.
- Sharing mapping work with non-mappers.

Professional orienteering mapmakers have a different approach to the GPS technology:

- A GPS device is used at an early stage of the ground survey to add more point and linear features; however, if we have a good base map, it is not very important whether we use GPS or not.
- A semi-professional GPS receiver is used in the initial stage of mapping. The mapmaker covers the terrain with the GPS receiver, recording anything that looks worthy, adding extra data: paths, walls, all kinds of point and line features. This enhanced map can be used as a base map for the ground survey.
- Real-time differential correction is used in the terrain with the orienteering map drawing software on a tablet PC. The main disadvantage of this hardware is not only the price, but also the lack of long time lightweight power supply.

Mapmakers can use other instruments like laser range finders and clinometers on the terrain, but using these devices is not widespread and these devices have not affected the process of mapmaking dramatically. Using these devices may increase the accuracy of terrain measurements or may speed up the time of measurement. These devices must be small enough and easily usable on the terrain even in difficult weather conditions (Zentai 2007).

2.7 Laser Scanning

One advantage of airborne laser scanning compared to classical stereophotography is that laser scanners are not dependent on the sun as a source of illumination (good aerial photos for orienteering maps can only be taken in certain weather and light conditions). If we use a precise digital terrain model, we can get more detailed relief information even in the case of dense vegetation. Because the technology can provide information not only about the illuminated top layers of the forest canopy, but signals from the surface can also be processed, users can extract valuable information for orienteering mapping: vegetation density/runnability. Technology could be ideal for orienteering maps, but it is still not easily available in every country or sometimes it is too expensive for orienteering map projects.

Advantages of laser airborne scanning for orienteering maps:

- Saving of time in the terrain because of its availability without interruption and its preciseness.
- Time-consuming work to find suitable photogrammetric pictures is not any more necessary.
- Significant reduced costs compared to photogrammetric base maps (if the raw data are already available).
- Combination with other georeferenced products (e.g. orthophotos) is easy.

The main risk of using laser airborne scanning in orienteering mapping is that the detailed contour relief of raw data could easily lead to an overcrowded and poorly generalized map image. Unfortunately, in certain countries where this data is freely available its use has changed the characteristic of orienteering competitions considerably: the competitors become slower, because they continuously want to identify all relief features that they see on the map.

3 Conclusion

One of the most important lessons that we have learnt from the example of the orienteering map is how the user requirements can influence the mapmaking process. If the maps are special, sometimes the users themselves have to learn the

cartographic techniques, because the cartographers are not able to make such maps for them. Orienteers have been trained for this job in the last fifty years and they flexibly adapt new tools and technologies.

Orienteering maps remained the only classic topographic products: they are field-checked and they are still printed on paper as required by their users.

References

Berglia K et al (1987) Orienteringsidretten i Norge gjennom 90 år. Norges Orienteringsforbund, Oslo

Spiess E (1972) International genormte topographische Karten für den Orientierungslauf. Int Yearb Cartogr 12:124–129

Zentai L (2007) New technologies in making orienteering maps. In: Abstracts of papers, XXIII international cartographic conference, Moscow, pp 343–344

Zentai L (2011) Legibility of orienteering maps: evolution and influences. Cartogr J 42:108–115. doi:10.1179/1743277411Y.0000000008

Map Projection Reconstruction of a Map by Mercator

Marina Rajaković, Ivka Kljajić and Miljenko Lapaine

Abstract The paper describes the beginning of research on Mercator's map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*. This map has many editions and shows a great part of present-day Croatia. Zagreb, the capital of Croatia, is shown in two different places on the map. Naturally, this raises the question of its accuracy. One of the first steps in the research process was finding the mathematical basis, i.e. the map projection used to create the map. The research results showed it was a trapezoidal projection, but there are no references as to where an appropriate derivation of the equations can be found. Therefore, on the basis of Mercator's own description, cited in this paper, the derivation of the equations for his trapezoidal projection is given. He used this map projection to produce the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*, as well as many other maps.

Keywords Mercator · Map projections · Trapezoidal projection

1 Introduction

In 2012, we celebrated the 500th anniversary of the birth of Gerard Mercator, one of the most famous cartographers of all time. It was an opportunity to remind us of his life and work. A symposium was held and an exhibition opened at the Faculty of Geodesy, University of Zagreb. For the purpose of the exhibition, the National and University Library in Zagreb (NUL) provided the Croatian Cartographic

M. Rajaković (✉) · I. Kljajić · M. Lapaine
University of Zagreb, Zagreb, Croatia
e-mail: mrajakovic@geof.hr

I. Kljajić
e-mail: ikljajic@geof.hr

M. Lapaine
e-mail: mlapaine@geof.hr

Society (CCS) with facsimiles of Mercator's maps kept in its Collection of Maps and Atlases. Some original maps are pages from atlases kept in the Collection, while others are single sheets taken from atlases. Facsimile text pages from the atlas, which are important because they describe the territory of today's Croatia, were obtained in addition to the facsimile maps.

Among the 30 or so facsimiles received from the NUL, the map entitled *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* is especially interesting for several reasons. First, it shows the territory of present-day Croatia. The second reason is reflected in the large number of editions of this map. For example, the NUL keeps nine different editions. In addition, the Novak Collection (Collectio Mappae Croatiae Prof. Dr. D. Novak), which was recently added to the NUL Collection of Maps and Atlases in Zagreb, includes seven different editions of the same map. Other editions of this map can be found in libraries and antiquarian shops around the world.

The second volume of Mercator's atlas appeared in 1589, entitled *Italiae, Sclavoniae, et Graeciae tabulae geographicae/Per Gerardum Mercatorem Illustrissimi.* It contains 22 maps, including the map *Sclavonia, Croatia, Bosnia cvm Dalmatiae parte* (Keuning 1947). Mercator's entire work was published in 1595 as *Atlas sive Cosmographicae meditationes de fabrica Mvndi et fabricati figvra*. The map *Sclavonia, Croatia, Bosnia cvm Dalmatiae parte* also featured in this edition.

Furthermore, Keuning (1947) noted that in the same year as Mercator's atlas was published (1595), 12 Latin editions of the atlas also appeared. The 13th and 14th releases were published in 1601 and 1603, respectively, while in 1598, Dutch, French and Italian editions, and in 1602, Spanish and Italian editions were published. The first Amsterdam edition of the *Atlas sive Cosmographicae meditationes de fabrica Mvndi et fabricati figvra* appeared in 1606 and the second in 1607.

Taking all this into account, one can conclude there are indeed many different editions of Mercator's map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*. But there is a particular reason for researching this map. The capital of Croatia, Zagreb, is represented on the map in two different places! The correct location of Zagreb is labelled with the city's German name, *Agram,* but there is also a town with the Latin name *Zagrabia* near Sisak. These were the three main reasons that led to the idea of researching the accuracy of the map. For the purpose of analysis, the sample of the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* from the Novak Collection, Sign. ZN-Z-XVII- MER-1630?-1, dated about 1630 (Fig. 1), was selected. This sample was previously published in the monograph *Pet stoljeća geografskih i pomorskih karata Hrvatske/Five Centuries of Maps and Charts of Croatia* (Novak et al. 2005).

The first steps in studying the map were taken by a doctoral student, M. Rajaković, in a seminar paper for the Faculty of Geodesy, University of Zagreb (Rajaković 2012). MapAnalyst software was used to analyze the map. The program was developed at the Institute for Cartography of the Federal Institute of Technology (*Eidgenössische Technische Hochschule—ETH*) in Zurich. It is completely free of charge and available on the website http://mapanalyst.cartography.ch/. The purpose of the program is visualization and studying the accuracy of old maps.
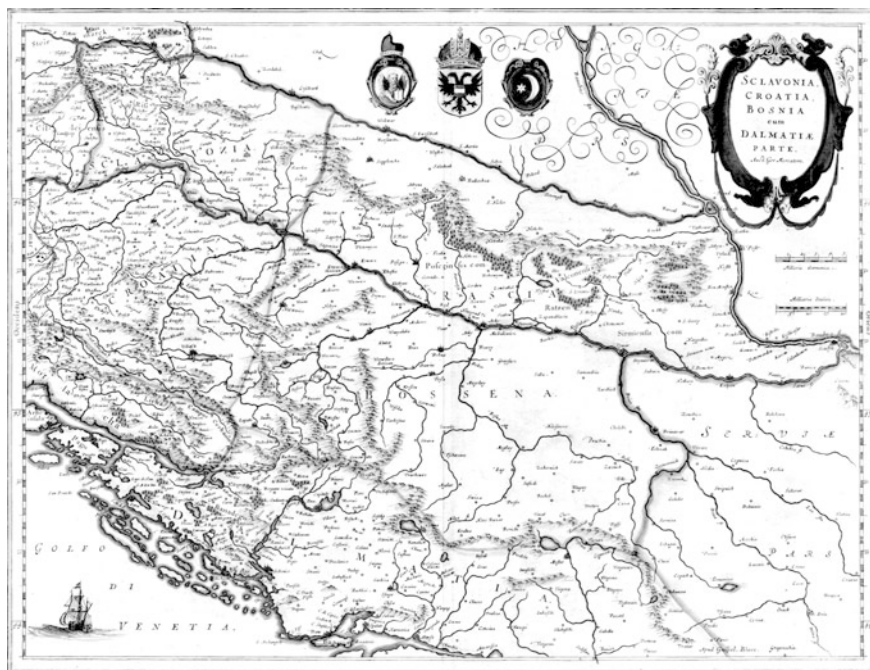
Fig. 1 Gerard Mercator, *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*, Collection Novak (Sign. ZN-Z-XVII- MER-1630?-1, dated about 1630)

Analysis began by searching for the same places on Mercator's map and the reference map. In order to compare results, the analysis was performed using two reference maps. The first was OpenStreetMap, which is supported within Map Analyst. The second was a general geographic map to the scale of 1:1000000 in a modified polyconic projection, with the ellipsoid adapted for the International Map of the World. This map was published by the Military-Geographic Institute in Belgrade in 1962 and updated in 1972. After choosing the type of transformation in MapAnalyst, visual representations of distortions, such as distortion grids, displacement vectors, rotation and scale isolines, were obtained. Numerical values of distortions were provided in the form of a report. In addition to MapAnalyst, graphics editing tools like GNU Image Manipulation Program (GIMP) and Inkscape were used to visualize distortions (Rajaković et al. 2012).

The results obtained were not sufficiently transparent, so it was decided to continue the research. The first task was self-evident: to determine the map projection Mercator used for his map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*.

2 Description of the Map Sclavonia, Croatia, Bosnia cum Dalmatiae parte

The map shows former areas of Slavonia (*Sclavonia*), Croatia (*Croatia*), Bosnia (*Bossena*), Rascia (*Rascia*) and parts of Dalmatia (*Dalmatia*), Serbia (*Servjæ*), present-day Lika and Slovenia. Part of Dalmatia with some islands is shown in the lower left corner. The River Drava (*Dravus fl*) and River Sava (*Savus fl*) are in the upper part of the map. They flow towards the Danube (*Danubius fl*), which is shown as far as the city of Belgrade (*Belgrado*). The River Drina (*Drina fl*), which divides the territories of Bosnia and Serbia, and the River Kupa (*Kulp fl*), which flows towards the Sava, are also drawn on the map. The map contains towns, rivers, mountains, forests, etc. with many toponyms in the territory of Croatia, Slavonia, Dalmatia, present-day Lika and Slovenia. Some towns are depicted with a signature of three towers in red. There are also towns with the same signature, but not in red. Some places are marked only by a small circle. The relief is drawn pictorially, like molehills. Mountains and hills and are shown in the same way. There is no information about their actual extent and height, just their location. It can be seen that the map was printed in black and white and coloured by hand afterwards.

A cartouche is located in the upper right corner containing the map title *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* and the name of the author, *Auct. Ger. Mercatore*. On the left side of the cartouche, in the middle upper part of the map, there are three coats of arms, those of the Republic of Venice (the lion of St. Mark), the Habsburg Monarchy and the Ottoman Empire (a crescent and star) (URL1). A ship which embellishes the map has been drawn in the Gulf of Venice (*Golfo di Venetia*), located in the lower left corner of the map (Rajaković 2012).

Two graphic scales are situated below the cartouche: the German mile (*Milliaria Germanica*) and Italian mile (*Milliaria Italica*). The German mile corresponds to 1/15 of one degree of latitude. The Italian mile corresponds to 1/60 of one degree of latitude. If the radius of a sphere is R = 6,370 km then it is:

$$\text{degree of latitude} = \frac{2R\pi}{360°} = 111.18 \text{ km}$$

$$\text{German mile} = \frac{\text{degree of latitude}}{15} = 7.41 \text{ km}$$

$$\text{Italian mile} = \frac{\text{degree of latitude}}{60} = 1.85 \text{ km}$$

$$\text{map scale (according to the German mile)} \approx 1 : 915000$$

$$\text{map scale (according to the Italian mile)} \approx 1 : 917000$$

The determination of map scale is based on measurement of each interval of the scale bar. After calculating the scale for each interval, the arithmetic mean of the scale was determined. According to the German mile, the map scale is 1:915000 and according to the Italian mile, it is 1:917000. It follows that the scale of the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* is approximately 1:916000.

One edition of the analyzed map was published in Mercator's famous work *Atlas sive cosmographicae meditationes de fabrica mundi et fabricati figura* (Atlas, or cosmographical reflections concerning the structure of the universe and the nature of the universe as created). This atlas is available in several libraries as well as in the NUL. However, we have this atlas on a CD in the Library of the Chair of Geoinformation and the Chair of Cartography at the Faculty of Geodesy of the University of Zagreb. Actually, it is a two CD release, published by Octavo, ISBN 1-891788-26-4. This digital edition enables us to access each page of the atlas using Adobe Acrobat Reader. There is also an English translation of all the texts. Images and maps can be zoomed. The whole atlas or individual pages can be printed out in black and white, or colour (Lapaine et al. 2008).

The atlas, which is also available on the Internet (URL2), contains the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*. It also includes Mercator's description of the map in Latin on page 406. The text in English reads:

Sclavonia, Croatia, Bosnia, and Part of Dalmatia

> Ecclesiastical administration from the *Provinciale Romanum*.
> The Archbishop of Zara or Zadar (which is now thought to be called Zaram) (39:23.44:33), to whom are subject the Bishops of: Veglia or Krk (38:19.45:10 in the map of Friuli); Arbe or Rab (38:49.44:56).
> The Archbishop of Spalato or Split (40:54.44:0), under whom are the Bishops of: Trau or Trogir (40:40.43:56); Tina (40:34.44:34); Sardona (40:8.44:20); Nona (39:16.44:42); Almisa (41:14.53:54); Sebenico or Sibenik (40:16.44:24); Farensis.
> The Archbishop of Ragusa or Dubrovnik (see the map of Greece and Macedonia, 42:52.42:50), under whom are the Bishops of: Stagno (42:25.43:12); Rossonensis or Bossononensis, i.e., Risine, I think (43:30.42:44); Trebina or Trebinje (42:58.43:8); Cataro or Kotor (43:40.42:36); Bacensis or Rosensis (42:16.42:30); Budva (43:34. 42:16).
> The Archbishop of Antivari or Bar (43:32.42:12 on the map of Macedonia), under whom are the Bishops of: Dulcigno or Ulcinj (43:54. 42:6); Drinasco (44:20.42:28); Scutari or Shkodër (44:20.42:24); Sardensis, Surtarensis or Acittarensis or Arbensis.
> The central meridian is 41:30, to which the rest tend in the ratio of the parallels 44:40 and 46:0.

This last sentence allows reconstruction of the map projection in which the map was made. But before going further in researching Mercator's map, let us see what can be concluded about the accuracy of Mercator's data. Some places, with their geographic coordinates in brackets, are mentioned in the above text written by Mercator. Naturally, the question arises as to their relationship to contemporary, modern coordinates. These coordinates were taken from *The twenty-first Century World Atlas* (Fernández y Fernández 1998) and their datum definition is not specified. In Mercator's time, the Earth was considered to be a sphere. For this purpose, Table 1 was created and a chart made (Fig. 2).

Table 1 Comparison of Mercator's and contemporary coordinates with the corrected geographic latitude of Mercator's Omiš

Places	Mercator's coordinates		Contemporary coordinates		$\Delta\varphi$ M–C	$\Delta\lambda$ M–C	Mercator's coordinates minus average difference		$\Delta\varphi$ M'–C	$\Delta\lambda$ M'–C
	φ_M	λ_M	φ_C	λ_C			$\varphi_{M'}$	$\lambda_{M'}$		
Zadar	44°33′	39°23′	44°07′	15°14′	26′	24°09′	44°05′	15°05′	−2′	−9′
Rab	44°56′	38°49′	44°45′	14°45′	11′	24°04′	44°28′	14°31′	−17′	−14′
Split	44°00′	40°54′	43°31′	16°26′	29′	24°28′	43°32′	16°36′	1′	10′
Trogir	43°56′	40°40′	43°32′	16°15′	24′	24°25′	43°28′	16°22′	−4′	7′
Knin	44°34′	40°34′	44°02′	16°11′	32′	24°23′	44°06′	16°16′	4′	5′
Skradin	44°20′	40°08′	43°49′	15°55′	31′	24°13′	43°52′	15°50′	3′	−5′
Nin	44°42′	39°16′	44°14′	15°11′	28′	24°05′	44°14′	14°58′	0′	−13′
Omiš	43°54′	41°14′	43°27′	16°41′	27′	24°33′	43°26′	16°56′	−1′	15′
Šibenik	44°24′	40°16′	43°44′	15°54′	40′	24°22′	43°56′	15°58′	12′	4′
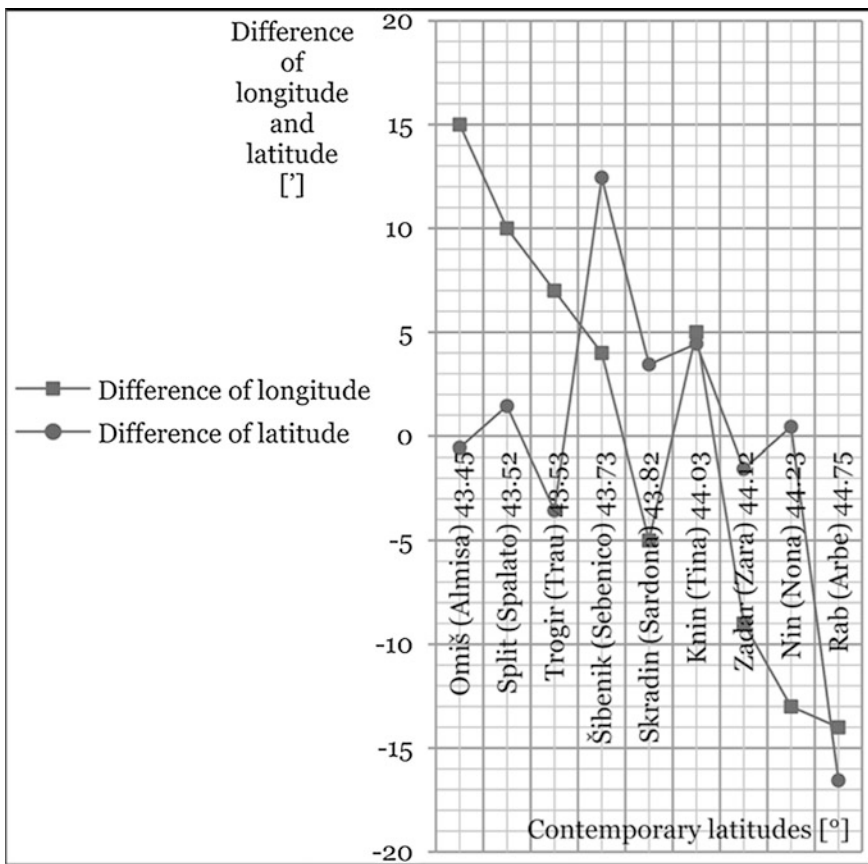
Fig. 2 Comparison of Mercator's and contemporary coordinates

Our research revealed that in Mercator's coordinates, there was at least one rough error or outlier (the latitude of Omiš). If we correct the latitude of Omiš to the real value (43°, Mercator has 53°), then we can get an average offset of $\varphi_0 = 0.46°$ S $= 28'$ S for Mercator's latitudes in the direction north–south. In other words, all Mercator's latitudes are about $28'$ greater than modern latitudes. Furthermore, we can estimate Mercator's prime meridian ($\lambda_0 = 24.30°$ W), because all longitudes are greater than modern longitudes by approximately this amount. Deviations remain after taking into account Mercator's systematic offset $\varphi_0 = 0.46°$ S $= 28'$ S and the estimation of the prime meridian as $\lambda_0 = 24.30°$ W. The source of these values could be incorporated into further research. The mean absolute value of the latitude deviation is $\Delta\varphi = 5'$ and the longitude deviation is $\Delta\lambda = 9'$. This means that Mercator knew the positions of the locations to an accuracy of approximately 10 km, which is approximately 1 cm on his map.

3 Discovering the Map Projection Step by Step

3.1 Data and Labels in the Frame of the Map

Mercator's map from the Novak Collection (Sign. ZN-Z-XVII- MER-1630?-1) contains written values of latitudes for every $2'$ on the right and left sides, in the map border. The Latin word *Occidens* (West) is written on the left side and *Oriens* (East) on the right side. According to the latitude, the area extends from 43°50' to 46°50' (values taken from Mercator's map). We first read the length of 10' intervals (from 43°50' to 44°00', from 44°00' to 44°10', etc.) on the left and right sides of the map's frame. This was done using the GIMP program. Then we compared them with each other (Fig. 3). It can be seen that deviations are less than a millimeter.

The longitude values are not marked or written on this map sample.

3.2 Graticule

It was difficult to determine the map projection because longitudes were not indicated. Therefore, we looked for other editions of the map with the same or similar content from approximately the same period, on which latitudes and longitudes were indicated.

The Collection of Maps and Atlases in the NUL in Zagreb includes the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*/per Gerardum Mercatorem (Sign. S-JZ-XVII-5), which is very similar to the map from the Novak Collection which we analyzed.
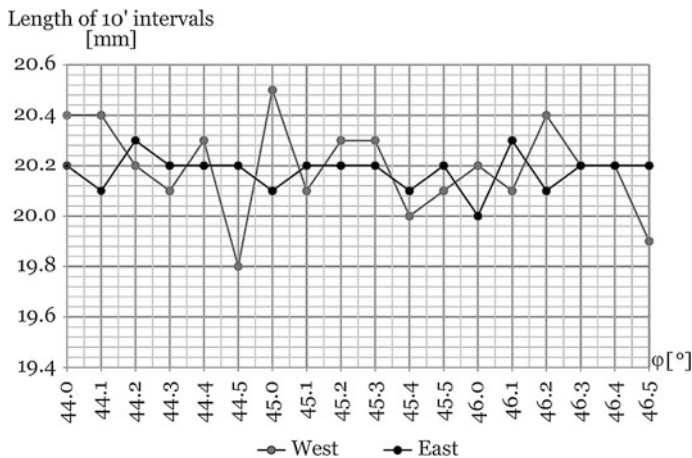
Fig. 3 Comparison of the length of 10′ intervals on both sides of the map frame

On this map, we connected equal values of latitudes marked on the left and right parts of the map with straight lines. We repeated this for the same values of longitudes (Fig. 4). The graticule in the projection indicates that a trapezoidal or related projection was applied.

3.3 Trapezoidal Projection

According to Keuning (1947), maps of larger territories were designed by Mercator in projections in the form of a trapezoid. However, Mercator did not take the parallels near the margin as the base, but parallels of latitude, which are removed from the central parallel as much as they are from the margin of the map. As a result of this, the old projection in the form of a trapezoid, which Donnus Nicolaus Germanus used to design his Ptolemaeus-edition map of 1482, was improved upon. In this projection, the central meridian was divided into equal parts and the parallels of latitude were drawn through the dividing points at right angles. The remaining meridians passed through the dividing points of the two parallels. Mercator explained his method of map projection construction in one short sentence. For example, for the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* he wrote, "Medius meridianus 41:30 reliqui ad hunc inclinatur ratione parallelorum 44:40 & 46:0".

Mercator's own words describing the projection can be found in his *Atlas sive cosmographicae meditationes de fabrica mundi et fabricati figura,* the content of pages 111 (right) and 112 (left), in Latin (URL2), or translated from Latin into English (URL3) in the chapter *Advice on Using the Maps.*
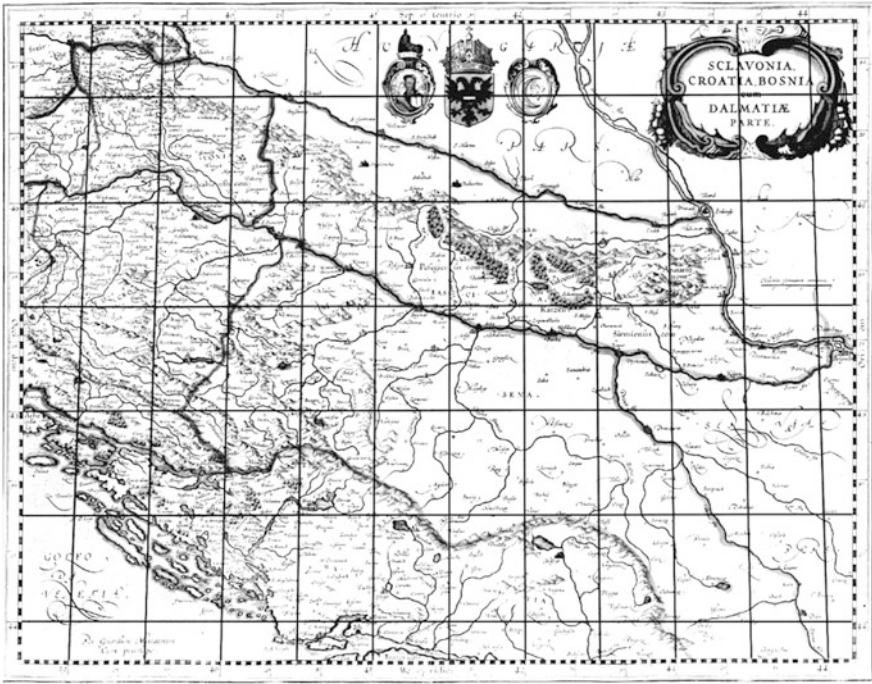
Fig. 4 A possible straight-line graticule on the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*/per Gerardum Mercatorem, NSK, Sign. S-JZ-XVII-5, dated about 1630

Mercator used a modified trapezoidal projection for many of his maps, starting in 1578. In order to reduce general deformation, standard parallels, or parallels along which linear distortions do not exist, are not on the edge of the map. They are placed approximately one quarter and three quarters along the distance between the latitude margins (Keuning 1947; Snyder 1993).

Beineke (2001) distinguishes normal trapezoidal projection, approximate solution of normal trapezoidal projection and free trapezoidal projection. But these three approaches do not describe Mercator's procedure as he described it, or as Keuning (1947) and Snyder (1993) interpreted it.

3.4 Derivation of Equations for Mercator's Trapezoidal Map Projection

We should emphasize that no textbook of map projections contains a derivation of equations of a general trapezoidal projection. Even the famous review of map projections by Grafarend and Krumm (2006) does no indication about using tra-peziodal projection except mentioning the Eckert II as an example of rectilinear

pseudocylindrical mapping. According to Frančula (2004), the shape of a normal cartographic grid of polyhedric projection does not fit any group of map projections. Beineke (2001) says this is probably because this "archaic" projection is a special case that does not fit into any of the usual theoretical classifications of map projections (azimuthal, cylindrical, conical, conformal, equivalent, equidistant) or others. Pseudocylindrical projections were probably omitted in Beineke's statement.

On the other hand, Beineke (2001) says that no modern maps were made on the basis of such an abstract and surreal cartographic grid. This is not true, because polyhedric projection was fairly widespread in the production of topographic maps in the last century.

However, the fact is that formulas which were used to create maps in polyhedric projection have a significantly different structure from others. Therefore it was considered that the projection could not be classified in any known group of usual projections.

Bugayevskiy and Snyder (1995) provide one type of formula for trapezoidal pseudocylindrical projection, known as the Müffling projection. Friedrich Carl Ferdinand Freiherr von Müffling was a Prussian general, military writer and geodesist (1775–1851). The first national military geodetic survey for the whole state of Prussia was performed under his direction. This laid the foundation for subsequent improvements connected with the names of Bessel, Bayer and Schreiber, which earned the Prussian state survey a world-wide reputation. Müffling's work was integral to the phase of radical change which geodesy underwent in the eighteenth and nineteenth centuries. During that period, the present definition of geodesy developed, as did international collaboration by many outstanding scientists and practitioners, so arc measurements and national geodetic surveys in Europe reached a high standard (Torge 2002). Müffling's projection was used in Germany, and later in Russia and the Soviet Union until 1928. The topographic map of the former Yugoslavia to the scales of 1:50000, 1:100000 and 1:200000 was made on the basis of topographic surveys (1920–1933) in polyhedric projection (Frančula 2004).

We will now show the derivation of formulas for Mercator's trapezoidal projection.

The following information is given:

Earth's sphere with radius R
Scale of projection 1:M
Longitude of central meridian λ_0
Latitudes of two standard parallels φ_1 and φ_2, presumption $\varphi_1 < \varphi_2$

Derivation:

First, we draw the x axis of the coordinate system, i.e. the image of the central meridian. Then we want all images of parallels to be mutually parallel lines perpendicular to the image of the central meridian, i.e. on the x axis (see Fig. 5). This means the equation for x should be

Fig. 5 Derivation of
Mercator's trapezoidal map
projection
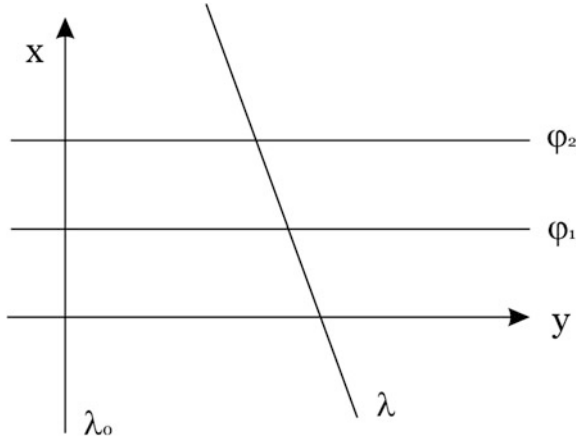
$$\mathbf{x = f(\boldsymbol{\varphi})}.$$

Presuming the projection is equidistant along the central meridian, we get

$$\mathbf{x = \frac{R}{M}\boldsymbol{\varphi}}.$$

Furthermore, we want the projection to have two standard parallels with corresponding latitudes φ_1 and φ_2, presuming $\varphi_1 < \varphi_2$. This means the point with coordinates

$$(\boldsymbol{\varphi_1}, \lambda - \lambda_0)$$

will be mapped into a point with coordinates

$$\left(\frac{\mathbf{R}}{\mathbf{M}}\boldsymbol{\varphi_1}, \frac{\mathbf{R}}{\mathbf{M}}(\lambda - \lambda_0)\mathbf{cos}\boldsymbol{\varphi_1}\right)$$

and the point with coordinates

$$(\boldsymbol{\varphi_2}, \lambda - \lambda_0)$$

will be mapped into a point with coordinates

$$\left(\frac{\mathbf{R}}{\mathbf{M}}\boldsymbol{\varphi_2}, \frac{\mathbf{R}}{\mathbf{M}}(\lambda - \lambda_0)\mathbf{cos}\boldsymbol{\varphi_2}\right)$$

Finally, the requirement is that the image of the point with coordinates

$$(\boldsymbol{\varphi}, \lambda - \lambda_0)$$

must belong to the straight line connecting images of the points

$$(\boldsymbol{\varphi_1}, \lambda - \lambda_0)\mathbf{and}(\boldsymbol{\varphi_2}, \lambda - \lambda_0),$$

i.e. in the projection

$$\left(\frac{R}{M}\varphi_1, \frac{R}{M}(\lambda - \lambda_0)\cos\varphi_1\right) \text{and} \left(\frac{R}{M}\varphi_2, \frac{R}{M}(\lambda - \lambda_0)\cos\varphi_2\right),$$

The equation of the straight line through these two points reads

$$y - \left(\frac{R}{M}(\lambda - \lambda_0)\cos\varphi_1\right) = \frac{\frac{R}{M}(\lambda - \lambda_0)\cos\varphi_2 - \frac{R}{M}(\lambda - \lambda_0)\cos\varphi_1}{\frac{R}{M}\varphi_2 - \frac{R}{M}\varphi_1}\left(x - \frac{R}{M}\varphi_1\right)$$

or, after minor editing

$$y = \frac{R}{M}(\lambda - \lambda_0)\frac{(\varphi - \varphi_1)\cos\varphi_2 - (\varphi - \varphi_2)\cos\varphi_1}{\varphi_2 - \varphi_1}.$$

Therefore, the final formulas for the Mercator trapezoidal projection are

$$x = \frac{R}{M}\varphi$$

$$y = \frac{R}{M}(\lambda - \lambda_0)\frac{(\varphi - \varphi_1)\cos\varphi_2 - (\varphi - \varphi_2)\cos\varphi_1}{\varphi_2 - \varphi_1}.$$

Based on the formulas derived, the graticule can be drawn in the projection plane. The graticule overlaps with Mercator's map (see Fig. 6).

4 Further Research

Future research might include:

- Further analysis of the accuracy of Mercator's map
- Determining the parameters of trapezoidal projections based on pairs of associated points
- Characteristic analysis of the trapezoidal projection (scale distortion, area distortion and angular distortion)
- Determining the inverse trapezoidal projection
- Comparing Mercator's trapezoidal projection with a polyhedric projection, Eckert I, Eckert II and Collignon's projections
- Defining general trapezoidal projection.
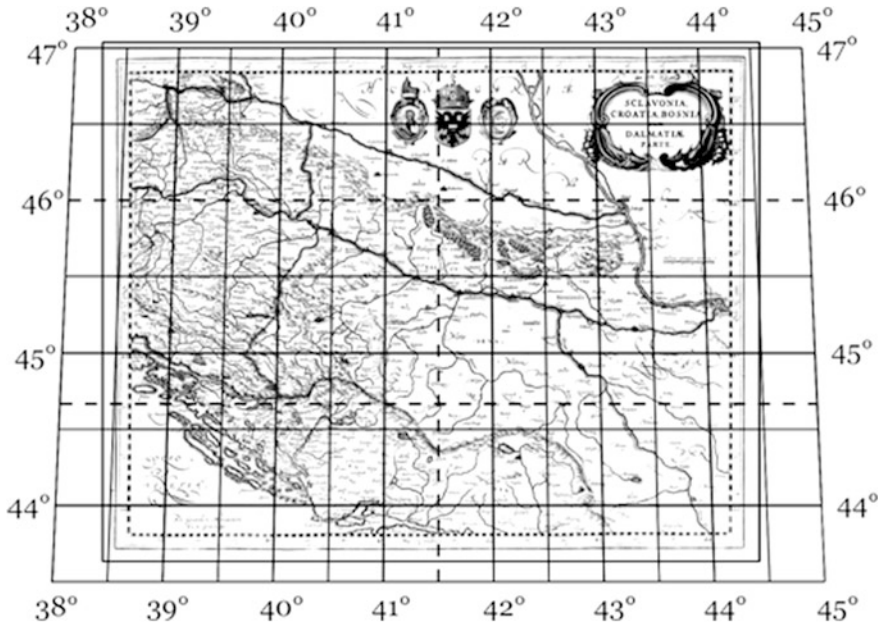
Fig. 6 The graticule in the projection plane, calculated using derived formulas for trapezoidal projection, and overlaid on Mercator's map. *Dashed lines* are standard parallels and the central meridian without linear distortions

5 Conclusion

Mercator's map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* was published in many editions and shows a great part of present-day Croatia. The map shows Zagreb, the capital of Croatia, in two different places, so this naturally raises the question of its general accuracy. To answer this question, we need to discover the mathematical basis of the map, i.e. find the map projection which was used to produce it. Our research results show that it was a trapezoidal projection, but as far as we know, there are no references as to where appropriate derivations of projection equations can be found. Therefore the derivations of the equations for Mercator's trapezoidal projection are given, based on Mercator's own description, as shown in this paper. According to Keuning (1955), Mercator used this projection not only in the production of the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*, but also for many other maps.

Further research should be carried out to determine the possible differences between theoretical map projections, as imagined and theoretically set by Mercator, and the real projection, which was incorporated as the basis of the map. It would then be possible to research the accuracy of other elements in the map.

References

Beineke D (2001) Verfahren zur Genauigkeitsanalyse für Altkarten. Schriftenreihe, Studinegang Geodäsie und Geoinformation Universität der Bundeswehr München. Heft 71

Bugayevskiy LM, Snyder JP (1995) Map projections: a reference manual. Taylor & Francis, London

Fernández y Fernández JA (ed.) (1998) The 21st century world atlas. Trident Press International, Naples

Frančula N (2004) Kartografske projekcije. Script, Faculty of Geodesy, University of Zagreb

Grafarend EW, Krumm FW (2006) Map projections. Springer, Berlin

Keuning J (1947) The history of an atlas: Mercator-Hondius. Imago Mundi 4:37–62

Keuning J (1955) The history of geographical map projections until 1600. Imago Mundi 12:1–24

Lapaine M, Perić O, Dubravčić I (2008) The origin of the term 'Atlas'/O porijeklu naziva atlas. Kartografija i Geoinformacije 9:42–53

Novak D, Lapaine M, Mlinarić D (eds) (2005) Pet stoljeća geografskih i pomorskih karata Hrvatske/Five centuries of maps and charts of Croatia. Školska knjiga, Zagreb

Rajaković M (2012) Točnost Mercatorove karte *Sclavonia, Croatia, Bosnia cum Dalmatiae parte*. Seminar paper, Faculty of Geodesy, University of Zagreb

Rajaković M, Kljajić I, Lapaine M (2012) The accuracy analysis of the map *Sclavonia, Croatia, Bosnia cum Dalmatiae parte* made by G. Mercator. In: Abstract, 8th cartography and geoinformation conference, September 28, 2012, Zagreb

Snyder JP (1993) Flattening the earth: two thousand years of map projections. The University of Chicago Press, Chicago

Torge W (2002) Müfflings geodätisches Wirken in der Umbruchepoche vom 18. zum 19. Jahrhundert. ZfV 2:97–108

URL1: Gothic Postal History. http://www.gothicstamps.com/php/viewitem.php?itemid=28366&germany=briefe&&PHPSESSID=33ca7b8a9b5898ef852cd37830bf8ac2#itempic1. Accessed 25 Oct 2012

URL2: Mercator Atlas, Latin version. http://mail.nysoclib.org/Mercator_Atlas/Octavo/McrAts3R.pdf. Accessed 12 Nov 2012

URL3: Mercator Atlas. http://mail.nysoclib.org/Mercator_Atlas/MCRATS.PDF. Accessed 9 Nov 2012

The Pole is Impracticable but *There* is a Land Northward: Austro–Hungarian Pole Expedition and Mapping of the Franz Joseph Land

Mirela Altić

Abstract Austro–Hungarian Pole Expedition (1871, 1872–1874) led by K. Weyprecht and J. Payer resulted with discovery of the Franz Joseph Land. On that occasion, first maps of Franz Joseph Land were created. The purpose of this work is to analyze Payer's and Weyprecht's original maps, the methods by which they were made and their influence on the subsequent development of polar cartography.

Keywords Polar cartography · Franz Joseph Land · Julius Payer

1 Introduction

The Austro–Hungarian North Pole expedition is richly represented in historical and geographical literature. The expedition's reports along with the related maps were periodically published in *Petermanns Geographische Mitteilungen.* The diary of Otto Krisch, one of the members of the expedition, published in 1875, was the first literary work that drew the attention of the general public to the expedition (Krisch 1875). The expedition was especially popularized by Julius Payer's book "*Die Österreich-Ungarische Nordpol Expedition in den Jahren 1872–1874*" in which he described the expedition and provided many accompanying illustrations

---

"Pole impracticable. No land to northward" is a legendary phrase wired in 1870 by captain Nares, the leader of the British Arctic Expedition which failed to achieve its fervently desired objective—to reach the North Pole.

M. Altić (✉)
Institute of Social Sciences, Zagreb, Croatia
e-mail: mirela.altic@zg.t-com.hr

(Payer 1876b). The book was very popular and the English edition was published in the same year (Payer 1876a).

The largest contribution to bringing the importance of the expedition to the general public in recent times was the exhibition marking the 100th anniversary of the discovery of Franz Joseph Land organised in 1973 in the Austrian National Library in Vienna. The publication of a comprehensive catalogue regarding the subject accompanied the exhibition. The exhibition and the catalogue presented the chronology of the expedition, living conditions of the crew, measuring instruments used on the expedition, personal histories and achievements of the main participants. However, Julius Payer's cartographic work and his maps remain under-appreciated and under-researched. The aim of this work is to reveal the process by which Payer made his maps, the methods of measurement he used and the degree to which his maps influenced the subsequent polar exploration in particular and the development of map-making in relation to polar areas in general.

2 Austro–Hungarian North Pole Expedition and its Role Model

Austro–Hungarian North Pole Expeditions were a short series of two mid-nineteenth century expeditions to the Arctic. The aim was to explore the North Polar Region and to promote the newly united, Habsburg and Hungarian Empire and its navy as a great power. Austro–Hungarian efforts in polar exploration were greatly influenced by German polar expeditions. Germany launched its first polar expedition in 1866. Karl Weyprecht, an experienced German explorer who would later play a crucial role in both Austro–Hungarian expeditions, was chosen to lead the expedition.[1] Julius Payer, a Czech explorer, took part in the second German Polar Expedition (1869–1870). He would later take part in both Austro–Hungarian expeditions in the capacity of cartographer and leader of field surveys and make the first maps of Franz Joseph Land.

The preparations for the Austro–Hungarian North Pole Expedition, made possible by the Geographic Society's enthusiasm and count Jozef Wiltczek's generous financial backing, proceeded under the watchful eye of August Petermann. Already in 1871 Austro–Hungarian Monarchy sent its first expedition ("Isbjörn" Expedition) led by Karl Weyprecht and Julius Payer. The aim of the expedition was to trace the route and leave supplies of food and equipment on the northern shores of the Novaya Zemlya archipelago for the main expedition that was to follow.

The second Austro–Hungarian Expedition, led by Karl Weyprecht and Julius Payer, headed out from the Norwegian port of Tromsö in June 1872. The name of the expedition's vessel was Tegettoff and the main objective of the enterprise was

---

[1] Weyprecht could not lead the expedition due to illness. Karl Koldewey led the expedition.

to explore the North East Passage. Having steamed past the western coast of Novaya Zemlya they continued due north. On the 30th of August 1873, at 79°43′ N–59°33′ E they sighted hitherto unknown land and they named the newly discovered region Franz Joseph Land (Payer 1876a). Having reached Cape Fligely at 82°05′ N and thereby setting the new world record they raised an Austro–Hungarian flag and then the crew decided to return to Europe. On the 20th of May 1874 the crew abandoned the icebound vessel and started on foot, with three small boats from the ship which they pulled on sledges, towards Novaya Zemlya. For almost three months people and dogs had to tug the boats on sledges, and then during next nine days they had to row until they finally met two Russian fishing boats near Novaya Zemlya. On the 3rd of November 1874 the Russians left them at the Norwegian port of Vardö. With that ended the 812 days long odyssey of the second Austro–Hungarian North Pole Expedition (Fig. 1).
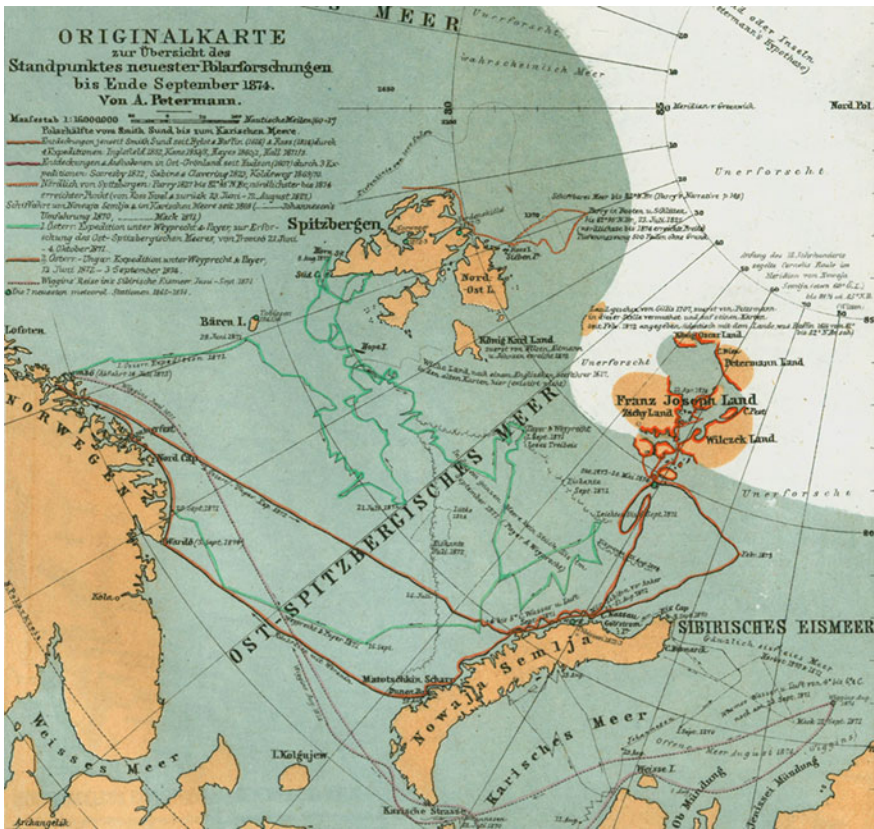
Fig. 1 Petermann's map showing the course of the Austro–Hungarian North Pole Expedition (*green*—first expedition, *red*—second expedition) (Petermann 1874)

3 Scientific Mission of the Austro–Hungarian Expedition and Measuring Instruments

What differentiates the Austro–Hungarian expedition from many previous polar expeditions is the fact that the leaders of the expedition insisted from the start that the collection of scientific data should be the primary concern of the expedition as opposed to claiming new lands for the crown or reaching the pole itself (Payer 1876a). In accordance with the scientific nature of the expedition many scientific measurements and observations were made in relation to meteorology, hydrography, geomagnetism, zoology and botany. Also, detailed geodetic-astronomical measurements were made, both for the purposes of updating existing and creating new maps (Fig. 2).

A theodolite was used to measure horizontal and vertical angles and the instrument was the main measuring instrument for geodetic surveys on land. The theodolite had been borrowed from the Nautical Academy in Trieste so as to keep the expenses as low as possible. In addition to the theodolite, one cross-prism (*Prismenkreis*), manufactured by Pistor and Martins and two sextants were also used for angular measurements. The cross-prism and sextants were used to determine the ship's position during the voyage. Two marine box chronometers (Barand 2/940 and Vorauer 67) and two pocket chronometers (Parkinson and

Auf dem Sonklar-Gletscher.

Fig. 2 Theodolite survey on Sonnklar Glacier (Payer 1876b)

Vorauer) were used for time measurements. Payer had used those same instruments on the German North Pole Expedition. An artificial horizon was used for inclination measurements. Also, a magnetic theodolite was used for measuring magnetic declination (Weyprecht 1878). The expedition achieved, through numerous measurements and observations, exceptionally valuable scientific results, including the cartographic representations of hitherto poorly explored or completely unknown Arctic areas. The achievement would not have been possible had the expedition not been well equipped and the crew professionally well prepared for the task.

4 Cartographers of the Austro–Hungarian Pole Expedition

In his capacity as the leader of land surveys Julius Payer was also the main cartographer of the expedition. He personally created almost all the maps made during the voyage. His extensive experience in mapping polar, as well as mountainous regions proved to be of crucial importance for the successful mapping of Franz Joseph Land.

Julius Ritter von Payer (2 September 1841, Teplice, Bohemia to 19 August 1915, Bled, Slovenia), received his education as soldier at Theresian Military Academy in Wiener Neustadt, where General Sonnklar was his teacher in geographic science. While serving with his regiment in Tyrol, he gained great celebrity as one of the most successful Alpine climbers. From 1864 to 1868 he explored the Adamello-Presanella Group and the Ortler Alps. His tours resulted in creating a detailed topographical map at a scale 1:56,000. Due to his achievements, Payer was transferred to the Austrian Military Cartographical Institute in Vienna.

Payer gained his first experience as polar explorer in the Second German North Polar Expedition (1869–1870). On that expedition Payer tried his hand for the first time at mapping polar areas. He created a whole array of vedutes, drawings, cross sections and one overview map of eastern Greenland. Payer's extensive military, exploring and cartographic experience made him the ideal candidate for leading, together with Carl Weyprecht, the first and second Austro–Hungarian expeditions when the monarchy started to make preparations for its own polar exploration effort. In 1875 he received a gold medal from the Royal Geographic Society for his achievements in the exploration of the Arctic and in 1876 he was elevated to the Austrian nobility. It is interesting to note that Payer left his military and cartographic career in 1874 and devoted his life to studying visual arts, an activity he pursued until he died at the age of 73.[2]

---

[2] Payer resigned from the army because of political maneuvers against him and his 'brother officers' doubts about his discovery and his sledge journeys. Österreichisches Biographisches Lexikon 1815–1950, Wien, 1978, 7: 374–375.

Even though somewhat overshadowed by Payer in a cartographic sense, Karl Weyprecht (8 September 1838, Darmstadt, Germany to 2 March 1881 Michelstadt, Germany), Payer's closest associate, nevertheless played an important part in map-making during the course of the expedition. In 1856, he joined the Austro–Hungarian Navy. During the following years he distinguished himself as a naval officer in numerous naval battles. He met August Petermann in 1865 during a meeting of the Geographic Society in Frankfurt.[3] The encounter altered the course of his life. Petermann aroused his interest in polar exploration and at Petermann's insistence he agreed to lead, together with Payer, both Austro–Hungarian expe-ditions. On account of his impressive nautical experience Weyprecht was tasked with the majority of astronomical observations which were crucial not only for safe navigation but also for accurate recognition of terrain features. After returning from the expedition he published many scientific works, including one concerning measuring base lines in relation to triangulation of Franz Joseph Land. Some astronomical and geodetic measurements were done by two other officers, Gustav Brosch (from Chomutov, Czechia) and Eduard Orel (from Nový Jičín, Moravia).

5 Weyprecht's Survey of the Novaya Zemlya

Weyprecht and Payer used the maps created by their predecessors for navigational purposes. Thanks to August Petermann who had been publishing in *Mitteilungen* reports and maps from all expeditions, Weyprecht and Payer had at their disposal the maps based on the most recent explorations of the Arctic region. During the first expedition, steaming across the aquatorium between Spitzbergen and Novaya Zemlya, the scientists and the crew, according to Payer (1876a), used Swedish maps on which certain errors were detected. Even though Payer did not specify the maps in question, we can safely assume that he referred to the maps of the Swedish expeditions led by Adolf Erik Nordenskiöld.

When an error was detected the Austrians tried to conduct their own surveys and accordingly modify the maps so that the errors would not appear on any future map copied from the ones they used. They detected a series of especially serious errors as they steamed along the western coast of Novaya Zemlya. Weyprecht and Payer used the maps that Russian explorer Friedrich Benjamin Lütke had made during the expedition of 1821–1824. Because Lütke's maps contained so many errors and because they did not contain any reliable information about the far north of Novaya Zemlya, Payer and Weyprecht decided to make their own map of that part of the coast, from Barents Island to the northernmost cape of Novaya Zemlya. In his book Payer (1876a) wrote: "The topography of the northern parts of Novaya Zemlya is complete confusion. The only survey which exists that of Lütke extends

---

[3] Allgemeine Deutsche Biographie, Leipzig, 1897, 42: 763–774.

no further than Cape Nassau. The maps of the Barentz Isles are frequently in contradiction with fact, and their correction is extremely desirable".

Between the 28th of August and the 3rd of October 1872 Weyprecht and Payer made many astronomical and geodetic measurements from the ship in order to make the map as accurate as possible. Payer would set the base line and then Weyprecht would measure the position and distance of the chosen points on the coast of Novaya Zemlya from the ship whose position had been astronomically measured the previous day. Altogether 15 points were observed (each of them from at least two positions of the ship). Based on that data they created their map, accurately representing the features of the coast (Petermann 1875).[4] Petermann followed up the effort by merging the newly acquired data with those acquired by surveys made during the Norwegian expedition of 1871 and created a new map of Novaya Zemlya. The quality of that map was extraordinary and represented a huge leap forward in terms of knowledge about the northern region of Novaya Zemlya (Fig. 3).
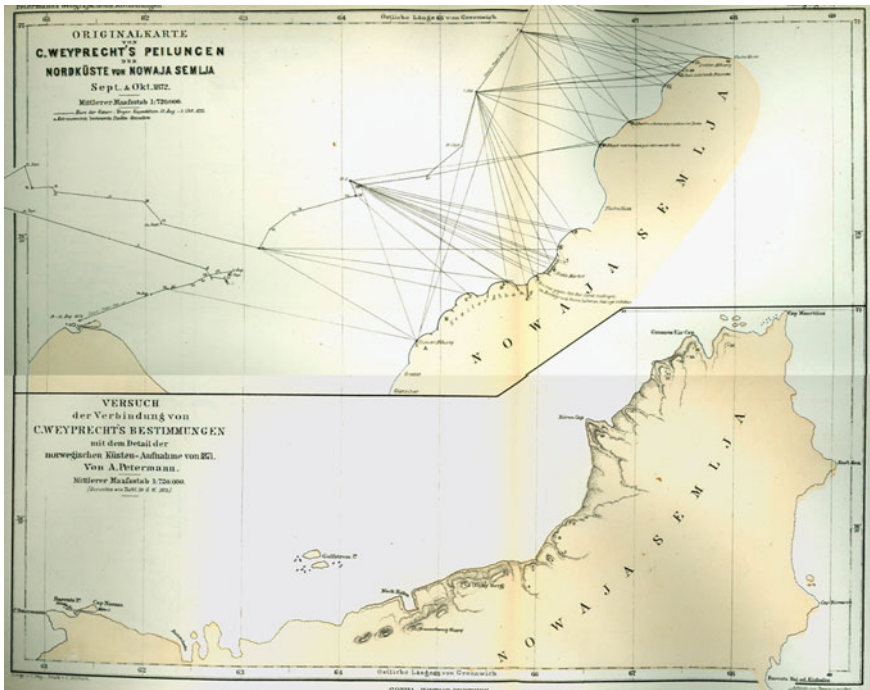
Fig. 3 Weyprecht's map of Novaya Zemlya (Petermann 1875)

---

[4] Originalkarte von C. Weyprecht's Peilungen der Nordküste von Nowaja Semlja: Sept. and Okt. 1872. Scale 1:720 000. Gotha: *Petermanns Geographische Mitteilungen,* 1875.

6 Discovery of Franz Joseph Land

Today we know that Franz Joseph Land comprises 191 ice covered islands. The Austro–Hungarian expedition discovered and explored the eastern part of the archipelago. The starting point of their exploratory efforts was the ship anchored at 79° 51′N 58° 56′E, to the south of an island they named Wilczek Island. From there a number of the crew went northwards on sledges, reaching the northernmost point of the archipelago, Cape Fligely on Prince Rudolf Island (Figs. 4 and 5).

They concluded that Franz Joseph Land consisted of two main land masses—the eastern landmass they called Wilczek Land and the western landmass Zichy Land. Between the landmasses stretched a passage they named Austria Sound. Austria Sound stretched northwards all the way to Prince Rudolph Land (island). To the south of Austria Sound they discovered a cluster of islands. The Tegetthoff lay at anchor there and the crew named some of the islands after the sponsors of their voyage and the others after famous polar explorers: Wilczek Island, Mac-Clintok Island, Hall Island, Salm Island, Lütke Island, Koldewey Island, Hochstetter Island. Pushing farther north they discovered a number of larger islands (Wiener Neustadt Island, Kane Island, Kuhn Island (lately Broch), Becker Island, Rainer Island and Hofman Island). Sometimes it was hard to distinguish where one island ended and another began because everything was covered with ice. For that reason Ronciere Island was mistakenly represented as a peninsular feature of Wilczek Land. Also Zichy Land, which Payer supposed was a compact land mass, was in fact a series of islands connected by ice (one of those islands was later named Payer Island). The discovery of Franz Joseph Land marked the beginning of the exploration and mapping of the archipelago. The cartographic representation of the archipelago, due to the climatic complexities of the region, has still not been brought to perfection.

7 Cartographic Survey of Franz Joseph Land

The main mapping effort of the expedition took place in Franz Joseph Land. The newly discovered land had to be charted and its main topographical features marked. The survey was conducted during three trips on sledges led by Payer in the spring of 1874 (10–16 March, 26–22 April, 29 April–3 May 1874). The following instruments were used: theodolite, cross-prism, sextant, compass, chronometer and aneroid barometer (Weyprecht 1878).

Payer created three maps of Franz Joseph Land on the expedition. He would first make a provisional drawing. Then he would improve on the drawing (second provisional map) and only then would he start creating the actual map. The process fairly accurately reflects the dynamics of the discovery of the new land. There are only the contours of Zichy Land and Wilczek Land and a few main islands in the

Fig. 4 Payer's first provisional map of the Franz Joseph Land (Petermann 1874)
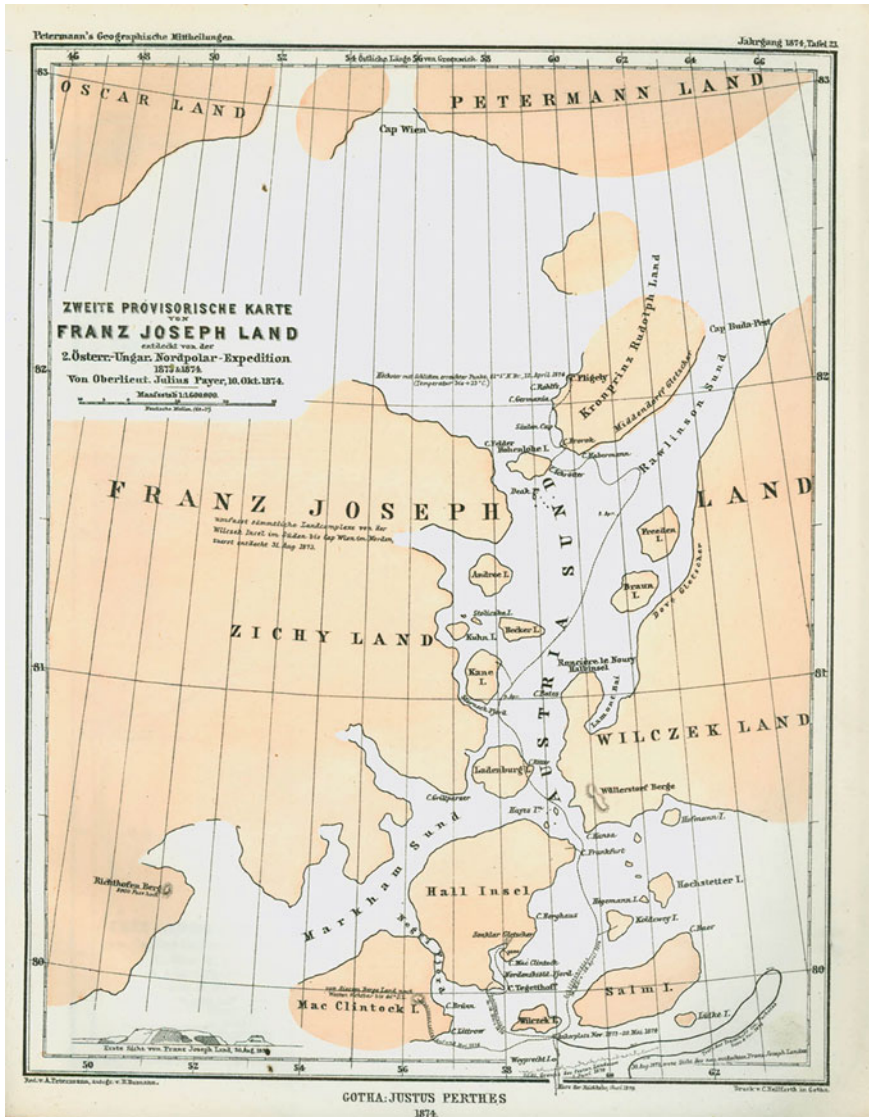
Fig. 5 Payer's second provisional map of the Franz Joseph Land (Petermann 1874)

initial drawing.[5] That drawing was based on visual observation and a few measurements conducted from a few small hills in the immediate vicinity of the ship. The second provisional map contains a comprehensively more detailed

---

[5] Provisorische Skizze von Franz Joseph Land/Oberlieutenant Julius Payer.—Scale 1:1.600000.—Gotha: *Petermanns Geographische Mitteilungen,* 1874 (Tafel 20).

representation of the coastal line and there is a whole array of islands not shown in the first drawing.[6] The final map was made on a scale of 1:1.000 000. The map accurately represents the coast and the island and also contains a whole array of topographic data such as costal terrain configuration represented by hachure lines and the height of individual peaks (in metres) and sea—depths, ice barrier (glaciated plateau), areas of unfrozen sea and positions of icebergs.[7] The map also contains a high number of toponyms given to various localities by members of the crew.[8] As mentioned, the map is a compilation of the cartographic surveys and all the other scientific data the expedition collected during its three voyages to Franz Joseph Land (Fig. 6).

The process of creating the map was made especially difficult by the position of the sun which was very low on the horizon, making the effort to determine the latitude very hard. In order to avoid adverse affects the mentioned conditions could have on accuracy of the survey, in addition to astronomical and geodetic determination of point coordinates, they included a trigonometric triangulation of the entire area of the archipelago (Payer 1876a). Suitable elevations in the terrain were chosen which were then connected in a triangle mesh. That meant, of course, that members of the expedition would have to climb many surrounding peaks. Altogether 15 trigonometric points were determined. The position of those points was determined by measuring the height of the sun, observation of stars (Northern Star, Orion) and planets (Jupiter, Mars). The theodolite and azimuth compass were used to measure vertical and horizontal angles between the trigonometric points and height above sea level was measured using the aneroid barometer (Petermann 1876). The triangular mesh was then connected to the base line whose position was determined through astronomical measurements done by Weyprecht. The northern point of the base line was a cross on the ship's mast while the southern point of the base line was marked by a cross on the ice crust not far from the ship. Between these two points the base line measured 2 170,8 metres and it was connected into the mesh of trigonometric points (Weyprecht 1878) (Fig. 7).

The mentioned maps were developed as field documents. Some smaller drawings were included in the ship's log (for example course charts) and larger maps were made as individual units. That especially relates to Payer's final map, the hand-drawn original of which is kept separately. Some smaller Payer's and Weyprecht's maps, published in *Petermanns Geographische Mitteilungen* are somewhat improved versions of the original. Petermann improved on the mathematical basis of the maps and their geographical nomenclature (Petermann 1874). In addition to the above, Petermann created two separate maps, based on Weyprecht's measurements of the ship's position and course charts, one map

---

[6] Zweite provisorische Karte von Franz Joseph Land/von Oberlieutenant Julius Payer.—Scale 1:1.600000.—Gotha: *Petermanns Geographische Mitteilungen,* 1874 (Tafel 23).

[7] Endgültige Karte von Franz Josef Land/aufgenommen von Julius Payer.—Scale 1:1.000 000.—Gotha: *Petermanns Geographische Mitteilungen,* 1875 (Tafel 11).

[8] The map contains the location of Otto Krisch's grave. Krisch was the only crew member who lost his life on the expedition. He was buried on the southern coast of Wilczek Island.

Fig. 6 Payer's final map of the Franz Joseph Land (Petermann 1876)
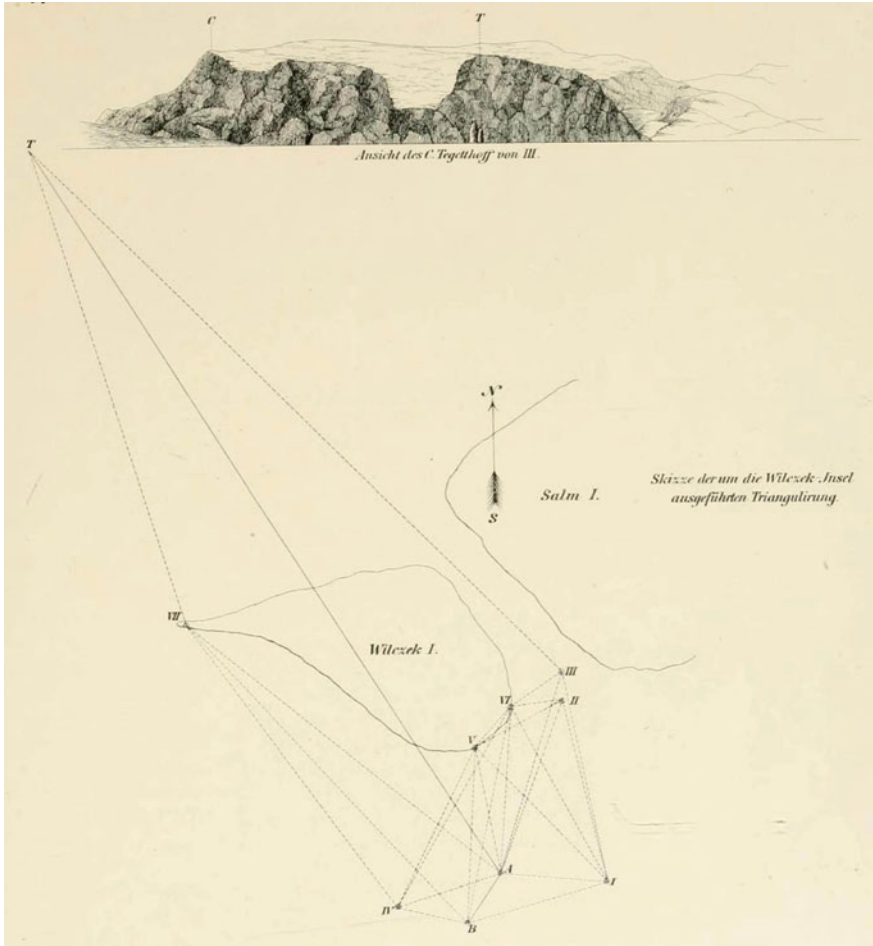
Fig. 7 Measuring of the base line near Wilczek Island (Weyprecht 1878)

showing, in minute detail, the course the Tegetthoff followed until it reached Franz Joseph Land[9] and the other the course of the return leg of the voyage.[10]

Besides the maps mentioned above, Payer created many longitudinal and transversal cross-sections of some of Franz Joseph Land's islands and many vedutes depicting Arctic landscape. Unlike the previously mentioned maps which

---

[9] Originalkarte der Eistrift der Österr.-Ungar. Expedition unter Weyprecht und Payer, 24. Aug. 1872 to 1. Nov. 1873./Von A. Petermann.—Scale 1:8 000 000.—Gotha: *Petermanns Geographische Mitteilungen,* 1875 (Tafel 12).

[10] Originalkarte der Rückreise der Österr.-Ungar. Expedition, Mai–August 1874. nach Weyprecht Beobachtungen/von A. Petermann.—Scale 1:1 900 000. Gotha: *Petermanns Geographische Mitteilungen,* 1877 (Tafel 5).
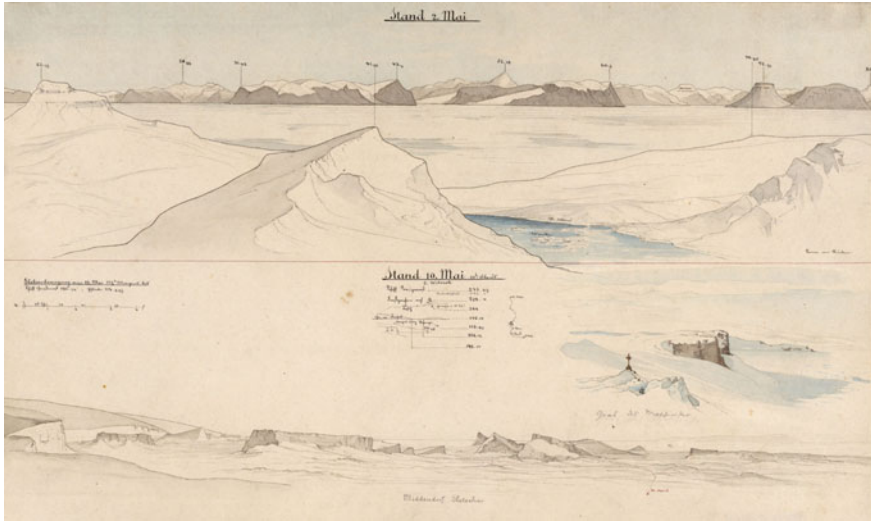
Fig. 8 One of the Payer's cross sections of Franz Joseph Land (Österreichische Nationalbibliothek, Wien)

were published as printed works by Petermann, Payer's cross sections and perspective representations of Franz Joseph Land remain only in the original hand-drawn form. The extraordinary accuracy and precision of Payer's works, as well as their aesthetic quality (drawn in pencil, painted with watercolours and ink), speak volumes about Payer as a dedicated scientist as well as a committed aesthetic devoted to painting. The latter inclination would play a major part in his later years (Fig. 8).

8 Concluding Remarks

With the discovery of Franz Joseph Land this Arctic archipelago was represented in maps for the first time. Payer's survey of the island was based on the most modern principles of cartography which, even though already widely used at the time, had never before been utilised with such precision in the harsh conditions of the Arctic climate. One example of those principles is the practice of combining the results garnered by astronomic and geodetic measurements. Knowledge about the topography of the newly discovered islands, as represented in top-view maps, was augmented by perspective representations, artistic vedutes and longitudinal and transversal cross-sections of the islands, which, at the time, was a new, all encompassing approach to exploration and mapping. Payer's multi-discipline approach, so clearly expressed in his maps also underlines the narrative of his book in which his description of Franz Joseph Land gives us a detailed account of the

geological, geomorphologic, hydrological, zoological, botanic and other features of the island, thereby creating a comprehensive picture about the geography of the newly discovered land. Such precision and consistency in ground reconnaissance set new standards in scientific research and mapping of polar areas.

The Austro–Hungarian North Pole Expedition made important steps in the context of the valorisation of the scientific data collected during the expedition. Its members published a whole string of scientific works presenting the results and methodology of their research. With that, they significantly contributed not only to the academic dissemination of knowledge but also to the viability of subsequent expeditions.[11] Karl Weyprecht distinguished himself in that regard by giving many lectures. His report, titled "*On Basic Principles of Arctic Exploration*", in which he explains the importance of establishing permanent weather stations, was presented in 1875 at the 48th German Convention of Scientists and Physicists in Graz. The report had a far reaching impact on the subsequent development of Arctic exploration, including cartographic reconnaissance. The very same ideas were again presented in 1879 at the 2nd International Congress of Meteorologists in Rome (Karl Weyprecht and Georg Neumayer discuss "setting up many stations in Arctic and Antarctic areas for the purpose of simultaneous meteorological and magnetic observation of the poles"), and finally in 1882/83 at the 1st International Polar Year event. In that sense we can safely conclude that the basic objective of the expedition, namely the collection of scientific data and the valorisation of the results with the aim of facilitating further scientific efforts, was completely met and that the scientific methodology used and achievements realised on the expedition had a groundbreaking impact on the progress of the subsequent mapping and exploration of polar areas.

References

Catalogue (1973) 100 Jahre Franz Josefs-Land: Zur Erinnerung an die Entdeckungsreise der Österreichisch-Ungarischen Nordpol-Expedition 1872–1874 unter Julius von Payer und Carl Weyprecht. Österreichische Nationalbibliothek, Wien

Krisch O (1875) Tagebuch des Nordpolfahrers Otto Krisch. Wallishauser'sche Verlagsbuchhandlung, Wien

Payer J (1876a) New Lands within the arctic circle: narrative of the discoveries of the Austrian ship "Tegetthoff" in the years 1872–1874. MacMillan & Co, London

Payer J (1876b) Die Österreich-Ungarische Nordpol Expedition in den Jahren 1872–1874. A. Hölder, Wien

Petermann A (1874) Die zweite Österreichisch-Ungarischen Nordpolar-Expedition unter Weyprecht und Payer, 1872/4. Petermanns Geogr Mitt 20:381–386

---

[11] The data collected during the Austro–Hungarian expedition significantly helped Nordenskiöld's expedition. He successfully negotiated the North East Passage in 1878//79. Also, the expedition of Fridtjof Nansen, which continued the exploration (and mapping) of Franz Joseph Land relied heavily and directly on Payer's and Weyprecht's results.

Petermann A (1875) Carl Weyprecht's Aufnahme der Nordküste von Nowaja Semlja, September und October 1871. Petermanns Geogr Mitt 21:393–394

Petermann A (1876) Die Entdeckung der Franz Josef-Landes durch die zweite Österreich-Ungarische Nordpol-Expedition, 1873 und 1874. Petermanns Geogr Mitt 22:201–209

Weyprecht C (1878) Astronomische und geodätische Bestimmungen der Österreichisch-Ungarischen arctischen Expedition 1872–1874. Denkschriften der kaiserlichen Akademie der Wissenschaften, Wien 35:47–68

Part II
Cartographic Tools

Enhancing the Locational Perception of Soft Classified Satellite Imagery Through Evaluation and Development of the Pixel Swapping Technique

Milad Niroumand Jadidi, Mahmoud Reza Sahebi and Mehdi Mokhtarzade

Abstract Spatial component is the key and most likely the first element of map making so that accurate spatial information improves the locational perception of map users. In this regard, soft classified satellite imagery conveys class proportions within pixels; however spatial distribution of the sub-pixels remains unknown. So, different visualization techniques (e.g. pie-chart representation of the proportions) are suggested to communicate the detailed land cover information. However, in each of which, the perception of actual spatial location of sub-pixels is definitely difficult for map users. Recently, the Super Resolution Mapping (SRM) techniques have been developed for optimization of the sub-pixels spatial arrangement based on the concepts of spatial dependency. These are relatively new methods which a comprehensive study on their performance and also their decisive parameters is a central issue for sub-pixel land cover mapping. In this research, the binary Pixel Swapping (PS) algorithm, as a prominent SRM algorithm, is developed for multivariate land cover mapping and the accuracy of the proposed method is evaluated in two procedures of independent and dependent of the soft classification error. Likewise, the impact of some parameters (e.g. zoom factor, neighborhood level and weighting function) is investigated on the efficiency of the algorithm. According to the results, the overall accuracy of the PS technique is extremely dependent on the accuracy of its input data (outputs of the soft classification). Furthermore, as a key result of this chapter, it is indicated that by increasing the zoom factor, the overall accuracy of the algorithm decreases. Also, the second level of neighborhood and inverse/square inverse distance functions has demonstrated the highest accuracies. Considering lower values than 5 for zoom factor, overall accuracy of the algorithm is determined higher than 90 % in procedure of optimizing the sub-pixels spatial arrangement.

M. Niroumand Jadidi (✉) · M. R. Sahebi · M. Mokhtarzade
Department of Remote Sensing Engineering, K. N. Toosi University of Technology, Tehran, Iran
e-mail: milad.niroumand@yahoo.com

1 Introduction

Maps of land cover could be considered as the most prominent cartographic products derived from satellite imagery. In this way, wide variety of classifiers has been developed which originally falls into hard type of classification methods (Thomas et al. 1987; Foody et al. 2005; Richards and Jia 2006). In these methods, each pixel is assigned to a single class which represents the most abundant material resident in that pixel area and the classified results are visualized as an image with several distinctive colors (Wessels et al. 2002; Marcal 2005). Therefore, they would not be able to provide detailed information within mixed pixels. The mixed pixels problem affects data acquired by satellite sensors of all spatial resolutions (Tatem et al. 2002). To solve the problem, soft classification techniques such as linear spectral unmixing (Foody and Cox 1994; Garcia-Haro et al. 1996), fuzzy c-means (Bezdek et al. 1984) and neural networks (Paola and Schowengerdt 1995; Atkinson et al. 1997) are developed to provide sub-pixel land cover information. The resultant output of the soft classifiers is a set of fractions, or abundances, that indicate the proportion of corresponding end-members within the pixels (Holben and Shimabukuru 1993; Foody and Cox 1994). However, there is not any information about the spatial arrangement of sub-pixels (Atkinson 2005; Villa et al. 2011). So, visualizing and communicating the results of soft classifiers is a challenging problem. There are several methods suggested to visualize this information. In a simple approach, fractional maps are being displayed as gray-scale side-by-side images. In other studies, linear combination of the colors assigned to each fractional image is suggested to generate the final map (Jacobson et al. 2007; Cai et al. 2011). Also, a different approach was presented by Cai et al. (2007) for visualizing fractional maps, in which a pie-chart layer is used to display the material composition information at the sub-pixel level. Though, the communication process is more complex in these kinds of representations. For example, linear combination of the colors creates a high diversity of color values that most probably disturbs the selective perception of users. Similarly, according to a user testing carried out by Cai et al. (2011), reading individual pie-chart results requires longer response time to retrieve detailed information. Moreover, none of the mentioned methods possess the spatial component of land cover information properly and do not allow map users to see distributional patterns of sub-pixels.

Relatively recently, the Super Resolution Mapping (SRM) techniques are introduced to overcome the locational perception problem of soft classifiers' output. SRM converts a soft classification result into a finer scale (sub-pixel level) hard classification map by optimizing spatial arrangement of the sub-pixels (Atkinson 2005; Kasetkasem et al. 2005; Atkinson 2009). Although several methods have

been developed for SRM such as genetic algorithms (Mertens et al. 2003), Hopfield neural networks (Tatem et al. 2002), simulated annealing and the PS (Atkinson 2005; Thornton et al. 2006; Niroumand et al. 2012), profound study on their accuracy and also the effect of various parameters on their performance is still under evaluation.

The Pixel Swapping (PS) algorithm is one of the methods of interest in recent years in SRM area (Niroumand et al. 2012; Thornton et al. 2006; Atkinson 2005) that optimizes the spatial arrangement of sub-pixels by maximizing the spatial dependency between neighboring sub-pixels. In this study, the binary PS algorithm (Atkinson 2005) is developed to multivariate sub-pixel land cover mapping and the overall accuracy of this algorithm is evaluated along with different influential factors on its efficiency (e.g. zoom factor, neighborhood level and weighting function). To avoid SRM input error resultant from output of the soft classifier, simulated fractional maps are used and the overall accuracy of the algorithm is calculated precisely. On the other hand, another accuracy assessment procedure is applied to evaluate the effect of the soft classification results (as SRM input) on the final super resolved map. Towards this end, fractional maps of the degraded image of the study area are estimated by the Linear Spectral Mixture Model (LSMM) and then have been super resolved by the PS algorithm.

The rest of this chapter is organized as follows. Section 2 outlines the theoretical and development procedure of PS technique. In Sect. 3, two different scenarios for accuracy assessment of the developed algorithm are being discussed. Furthermore, influential parameters on the performance of the algorithm are addressed and evaluated comprehensively in Sect. 4. Section 5 presents the results of empirical implementation. The chapter is finalized in Sect. 6 presenting some conclusions and outlooks.

2 Development of PS Algorithm

In recent years, more attention has been focused on PS algorithm (Niroumand et al. 2012; Atkinson 2009). The basic algorithm has been applied on binary target detection issues (Atkinson 2005) and also for sub-pixel mapping of specific land cover features (Thornton et al. 2006; Foody et al. 2005). The algorithm is founded based on the first law of geography (Tobler 1970) indicating that the near things are more related than distant things. So, the basic idea is to maximize the spatial dependency between neighboring sub-pixels under the constraint that the original pixel proportions were maintained (Atkinson 2004). In this regard, after initial allocation of the sub-pixels, their labels only swap within the corresponding pixel. Swaps should be organized in a way that causes increasing the spatial dependency between neighboring sub-pixels. In the present chapter, the PS technique is developed for multivariate land cover mapping in an iterative procedure by applying the basic binary algorithm for each of desired land cover classes. In the rest of the section, the developed algorithm is being described in details.

Considering a specific Zoom Factor (*ZF*), each pixel is divided into $ZF \times ZF$ sub-pixels. Thus, the number of sub-pixels related to each one of classes could be calculated according to Eq. (1).

$$NSP_k(x, y) = \left[F_k(x, y) \times ZF^2 \right]. \tag{1}$$

In the above equation, NSP_k[1] (x, y) represents the number of sub-pixels per *k*th class in the pixel coordinate (x, y) and $F_k(x, y)$ is the fractional information per *k*th class in the same pixel coordinate acquired from soft classification.

Then, the calculated numbers of sub-pixel proportions are allocated randomly within each pixel. Once allocated, only the spatial arrangement of the sub-pixels can vary within the corresponding pixel to keep the number of sub-pixels allocated within each pixel fixed. In order to achieve the optimal arrangement of sub-pixels, the swapping procedure attempts to maximize their spatial dependence so that sub-pixels close together to be more alike than those that are further apart with respect to land cover types. In this regard, for every sub-pixel, the attractiveness of the location is predicted as a distance weighted function of its neighbors (Eq. 2).

$$A_k(u, v) = \sum_{j=1}^{n} W(x_j, y_j) \times F_k(x_j, y_j) \tag{2}$$

$$W(x_j, y_j) = 1/d(x_j, y_j), \quad d(x_j, y_j) = dist\left[(u, v), (x_j, y_j)\right] \tag{3}$$

where $A_k(u, v)$ represents the amount of attractiveness per *k*th class in the pixel coordinate (x, y) and $F_k(x_j, y_j)$ is abundance of the same class in the *j*th neighboring pixel. Also *n* stands for the number of neighboring pixels involved in calculations and $W(x_j, y_j)$ is weighting function (Eq. 3) that has an inverse relation with the distance between the pixel location for which the attractiveness is desired and the location of the *j*th neighbor, $d(x_j, y_j)$.

Atkinson (2005) proposed an iterative swapping process to distribute sub-pixels in locations with probably most attractiveness values. The original algorithm is suggested for binary target detection purposes; however this research attempts to develop it for multivariate land cover sub-pixel classification. Achieving this, the main swapping concept is repeated for each of the land cover classes. Considering the attractiveness values of the *k*th class in pixel coordinate (x, y), the least attractive sub-pixel location currently allocated to desired class is stored as (u_1, v_1).

$$(u_1, v_1) = Arg \min_{u,v} \{A_k(u, v) \mid (u, v) \in k\} \tag{4}$$

Similarly, the most attractive location currently allocated to a different class than desired class ($k' \neq k$) is also stored as (u_2, v_2).

$$(u_2, v_2) = Arg \max\left\{A_k(u, v) \mid (u, v) \in k', \ k' \neq k\right\}. \tag{5}$$

---

[1] Number of Sub-Pixels (NSP).

If the attractiveness of the least attractive location, $A_k(u_1, v_1)$, is less than that of the most attractive location, $A_k(u_2, v_2)$, then the classes are swapped for the pixel in question (Eq. 6). If it is more attractive, no change is made.

$$if \quad A_k(u_1, v_1) < A_k(u_2, v_2) \qquad then \qquad (u_1, v_1) \in k' \quad and \quad (u_2, v_2) \in k. \quad (6)$$

After completing the above steps for all image pixels, the process is repeated for other classes as well. Thus, the first iteration of the algorithm is done. The whole process is then repeated until a specified number of iterations have taken place, or the swapping algorithm does not make any change in the sub-pixel locations.

3 Accuracy Assessment

As mentioned earlier, more investigations are eligible in order to assess the accuracy and overall performance of SRM algorithms. In this research, two different approaches of dependent and independent of the soft classification error have been attempted to explore the overall accuracy of the multivariate PS algorithm.

3.1 Independent from Soft Classification Error

The final accuracy of SRM algorithms is influenced by the accuracy of soft classification results. So, to have a precise and detailed evaluation of SRM methods, it is inevitable to examine their accuracy only in the optimization process of sub-pixels arrangement independent of the input data error. For this purpose, simulated proportions are used according to a suggestion provided by Atkinson (2009). In this way, firstly the original satellite image is labeled as land cover classes by means of a hard classification method (e.g. maximum likelihood algorithm). Then by moving a $ZF \times ZF$ window across the target hard classified image and so counting the number of each land cover class located in each window, a coarse spatial resolution proportions image has been produced. Thus, simulating zero error in the land cover proportions provides the possibility to assess the optimization of sub-pixels arrangement through PS process more exactly by considering the hard classified map as a target (Fig. 1).

According to the fact that the proportions image is simulated based on a real satellite image, all the characteristics of class diversity and distribution have an actual condition.
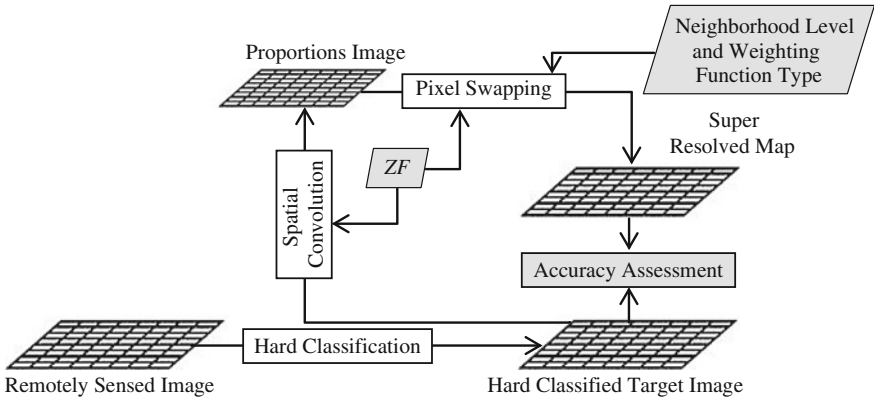
Fig. 1 Accuracy assessment of the multivariate PS algorithm independent of the soft classification error, adopted from Atkinson (2009)

3.2 Dependent to the Soft Classification Error

The accuracy assessment of SRM algorithms in the presence of soft classification error requires sub-pixel reference data which is mostly hard to access. An alternative to the above testing scenario is to take a remotely sensed image and degrade it from its original spatial resolution to a coarser one (Atkinson 2009). Then the original image can be classified to land cover classes providing a target image for accuracy assessment. In this chapter, the LSMM is used to predict the class proportions as PS input data (Fig. 2).
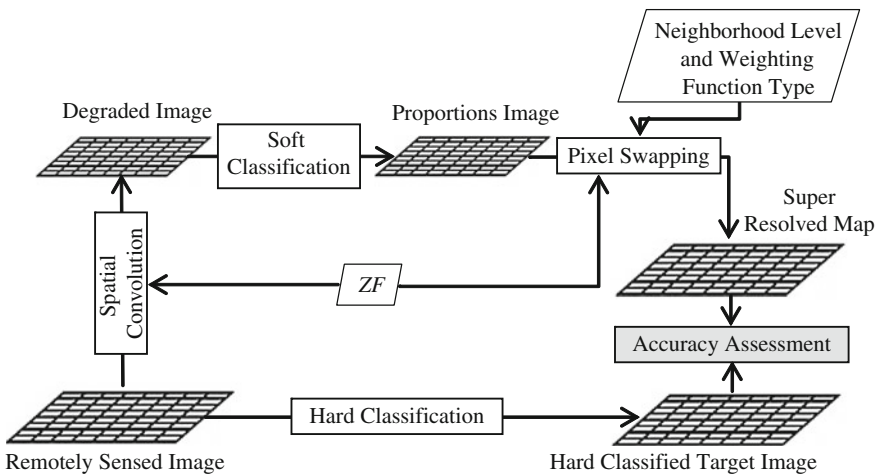
Fig. 2 Accuracy assessment of the multivariate PS algorithm dependent of the soft classification error, adopted from Atkinson (2009)

4 Influential Parameters on the PS Performance

In addition to overall accuracy of the PS algorithm, parameters affecting the performance of this algorithm (highlighted in the Fig. 1) are also addressed.

4.1 Zoom Factor (ZF)

This parameter is a main factor affecting the accuracy of the PS algorithm. So, several different values of *ZF* are considered through accuracy assessment of the algorithm in the implementation stage.

4.2 Neighborhood Level

As mentioned in Sect. 2, the optimization of sub-pixels' spatial arrangement requires attractiveness calculation of each sub-pixel location with respect to each of the desired classes based on the neighboring pixels. In this regard, neighborhoods up to fifth level (e.g. eight pixels in the second neighborhood level) are studied on overall accuracy of the super resolved land cover map (Fig. 3).

4.3 Weighting Function

The geometric distance between sub-pixel locations and their neighboring pixels forms the structure of weighting function which they have an inverse relation together. Different weighting functions such as the inverse distance (Eq. 3), the inverse square of distance (Eq. 7) and exponential function (Eq. 8) are used and their impact on overall accuracy of the PS algorithm has been evaluated. In the Eq. (8), "*a*" represents the non-linear parameter of the exponential function.

$$W(xj, yj) = 1/d(xj, yj)^2 \tag{7}$$

$$W(xj, yj) = \exp\left(\frac{-d(xj, yj)}{a}\right). \tag{8}$$

Fig. 3 First to fifth neighborhood levels in order to calculate the sub-pixels attractiveness

5 Implementation and Results

5.1 Study Area and Data Used

The study area is located in West-Azerbaijan province in Iran which is covered by four main land cover classes including water, pasture, garden and bare earth. A 900×900 pixel image of the Landsat *ETM+* sensor is used to sub-pixel land cover mapping (Fig. 4).

5.2 Empirical Results

First, the testing scenario independent of soft classification error has been applied on the study area to explore the proficiency of developed multivariate PS algorithm along with the impact of influential parameters on its overall accuracy. The hard classified target image, random initial allocation of sub-pixels and optimized pixel swapped map are shown in Fig. 5.

Five different neighborhood levels are examined in a range of zoom factors considering an inverse distance weighting function (see Fig. 6 and Table 1).

According to the Table 1, increasing the *ZF* has been reduced the overall accuracy of the PS algorithm. Furthermore, the second neighborhood level demonstrated the most overall accuracy (Fig. 7).

On the other hand, different weighting functions are tested by considering a constant zoom factor ($ZF = 3$) for first to fifth neighborhood levels (Table 2).
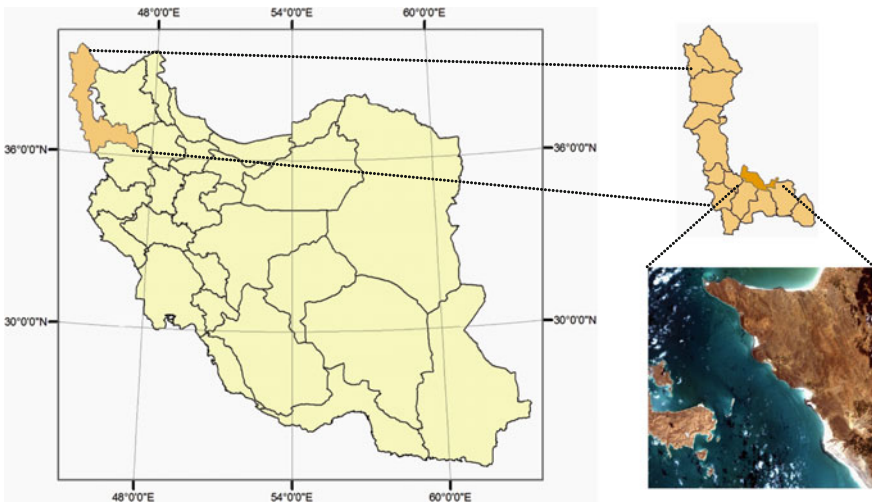
Fig. 4 The study area and Landsat ETM+ image
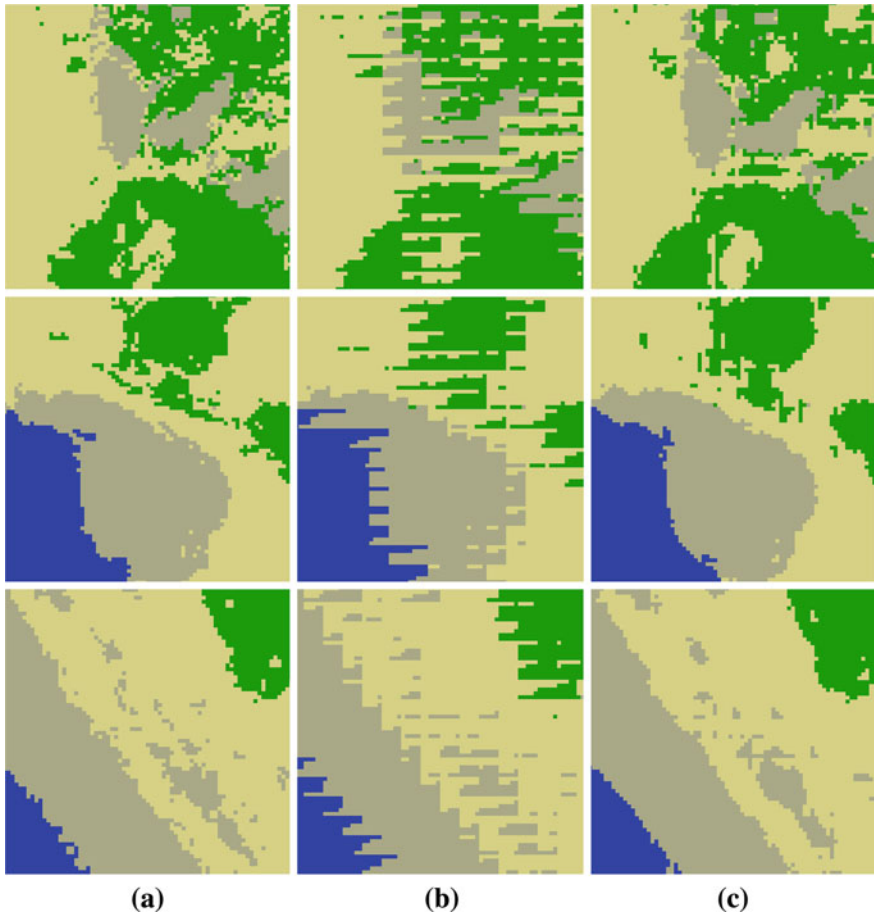
Fig. 5 Testing of multivariate PS algorithm independent of the soft classification error, Column **a** hard classified target images, Column **b** initially allocated sub-pixels, Column **c** optimized land cover maps ($ZF = 3$)

According to the results, the inverse/squared inverse distance functions would have a higher accuracy rather than exponential function. It should be noted that the optimum value for the nonlinear parameter of the exponential function is considered as $a = 5$.

As mentioned in Sect. 3.2, to investigate the whole process of sub-pixel land cover mapping dependent to the soft classification error, the original image is degraded by applying various ZF values and then LSMM is used to estimate the class proportions. Subsequently the sub-pixels are initially allocated and finally optimized by the PS algorithm. Then, hard classified map of the original image is used as a target for accuracy assessment. According to the Table 3, as expected, by entering the SRM input data error related to the soft classification process, the
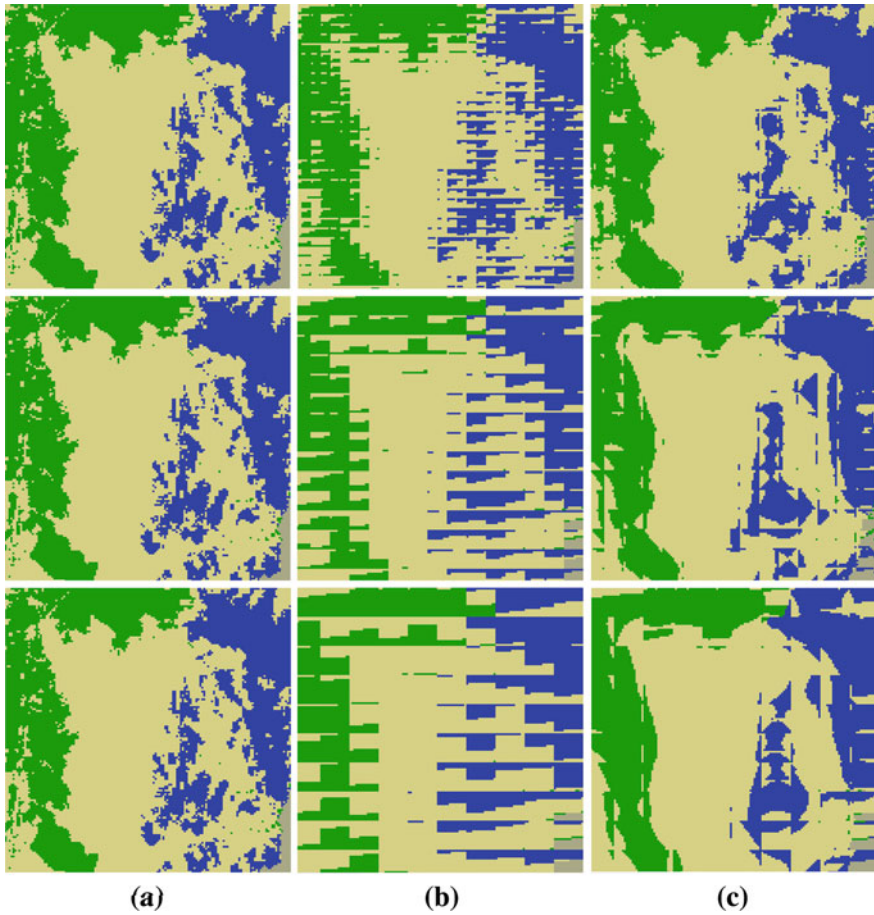
Fig. 6 The evaluation of the *ZF* parameter on overall accuracy of the multivariate PS algorithm, Column **a** hard classified target images, Column **b** initially allocated sub-pixels, Column **c** optimized land cover maps (*ZF* increases in each row as 3, 5 and 6 respectively)

Table 1 Overall accuracy of the multivariate PS algorithm independent of the soft classification error

Neighborhood level	Zoom factor (ZF)								
	2	3	4	5	6	9	10	12	15
I	94.64	93.31	90.04	87.34	84.93	79.94	78.41	74.86	70.87
II	94.89	93.64	90.26	87.51	85.11	80.12	78.62	74.92	70.93
III	94.57	93.27	89.92	87.32	84.76	79.83	78.37	74.62	70.76
IV	94.21	93.12	89.76	86.82	84.49	79.52	78.12	74.31	70.54
V	92.02	92.89	89.64	86.71	84.35	79.44	76.92	74.06	70.22
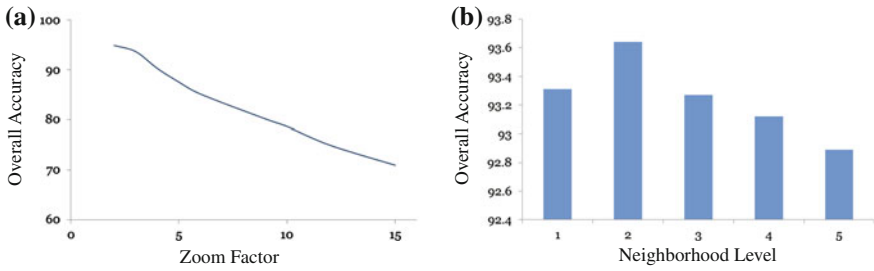
Fig. 7 The impact of the **a** *ZF* and **b** neighborhood level on overall accuracy of the multivariate PS algorithm

Table 2 Overall accuracy of the multivariate PS algorithm regarding three different weighting functions, $ZF = 3$

Neighborhood level	Weighting function		
	d^{-1}	d^{-2}	$\exp(-d/a)$
I	93.31	93.26	93.18
II	93.64	93.51	93.21
III	93.32	93.45	93.06
IV	93.12	93.31	92.91
V	92.89	93.07	92.64

Table 3 Overall accuracy of the multivariate PS algorithm dependent of the soft classification error

Neighborhood level	Zoom factor (*ZF*)								
	2	3	4	5	6	9	10	12	15
I	88.45	87.58	86.12	83.91	81.05	76.67	74.86	70.34	65.13
II	88.64	87.72	86.21	84.07	81.17	76.89	74.97	70.44	65.28
III	88.31	87.46	86.02	83.83	80.96	76.54	74.67	70.23	65.02
IV	88.22	87.31	85.88	83.72	80.89	76.41	74.59	70.11	64.85
V	88.14	87.23	85.75	83.64	80.73	76.33	74.48	70.02	64.76

accuracy of the PS algorithm is reduced considerably. The results of soft classifiers most likely contain some noises which lead to a small number of sub-pixels in the image. As illustrated in Fig. 8, these partial proportions have suppressed the sub-pixel land cover map. The final sub-pixel land cover map of the study area is given in Fig. 9.
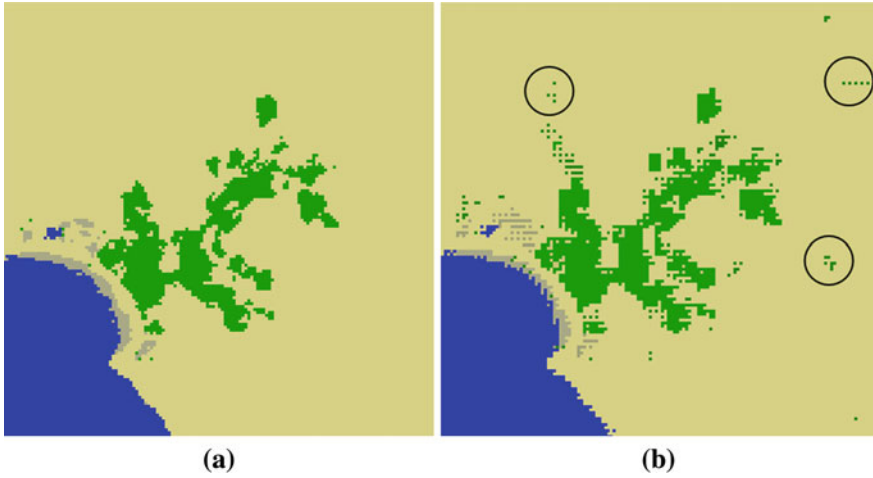
Fig. 8 The influence of soft classification error on the final sub-pixel land cover map; some partial noises are highlighted by *circles*, (**a**) hard classified target image, (**b**) optimized land cover map
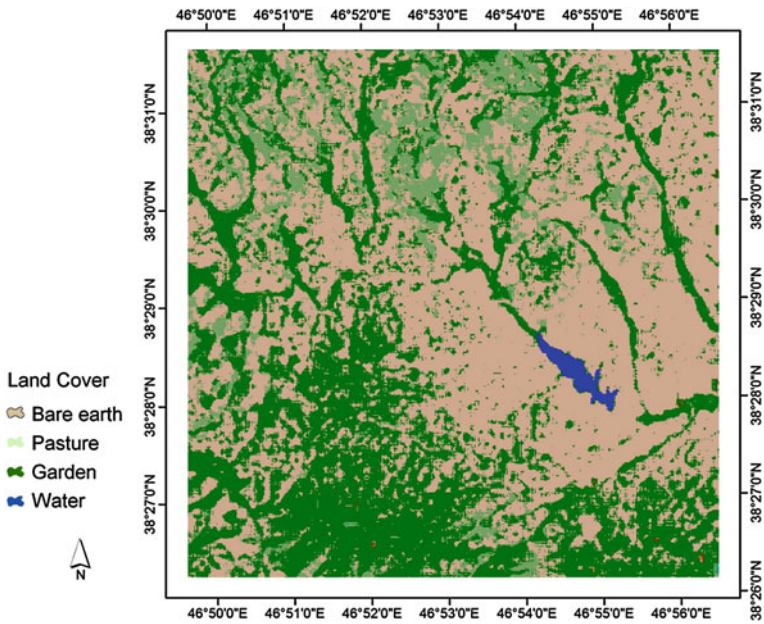
Fig. 9 Sub-pixel land cover map of the study area produced by the multivariate PS algorithm

6 Discussion and Conclusions

Soft classification of the satellite imagery is indispensable for accurate land cover information due to mixed pixels problem. However, these classification methods individually do not reveal the spatial component of the sub-pixels which is essential and basic element of map making. Some visualization techniques are previously presented for transferring the soft classification information to the users with no success to represent such information as a land cover map. In this regard, the SRM techniques have opened up a new horizon to sub-pixel land cover mapping using the soft classification outputs. In this chapter, the PS algorithm addressed comprehensively as the most interesting algorithm in this area. This algorithm is a spatial optimization technique which allocates randomly class labels to sub-pixels initially and then in an iterative procedure swaps the sub-pixels within each pixel in such a way to maximize the spatial dependence between neighboring sub-pixels. In this research, the basic binary PS algorithm is extended to the multivariate land cover mapping purposes. To study the accuracy of the algorithm, two testing scenarios are examined. In this way, simulated proportions are used to explore and test the proposed optimization algorithm independent of uncertainties related to the soft classification process. On the other hand, whole procedure of the SRM dependent to the soft classification error is tested using the degraded image and super resolving it to its original image size by the PS algorithm. With regard to these assessment techniques, the overall accuracy of the algorithm and the effect of parameters such as ZF, the neighborhood level and the weighting function are evaluated.

According to the results, by increasing the ZF, the spatial arrangement of sub-pixels will face with a higher degree of uncertainty and the overall accuracy of the algorithm will be decreased. However, a trade-off between the spatial resolution and overall accuracy of the PS classification is needed to be applied. The most accurate result of the algorithm independent of the soft classification error obtained by assuming $ZF = 2$ (15 m spatial resolution for Landsat data) and in the second neighborhood level as 94.89 %. By considering the same parameters in the case of dependent to the soft classification error, the overall accuracy decreased to 88.64 %. Moreover, it is concluded that the PS algorithm works more accurately in the second neighborhood level and it depends less on higher levels. It is for this reason that the third and higher neighborhood levels are located in a nonadjacent pixel layer to the desired pixel location that less similarity and dependency could be expected among them. Also, the resultant land cover maps using the inverse/square inverse distance weighting functions are relatively accurate than that of predicted by the exponential function. For instance, the PS algorithm using the inverse distance function shows 0.43 % (equal to 3483 pixels or 313 ha) higher overall accuracy relative to using exponential function in the second neighborhood level.

The testing scenario independent of the soft classification error represents the exact accuracy of the PS algorithm through spatial optimization of the sub-pixels influencing no source of error. However, the second testing scenario includes a source of uncertainty related to the reference image which is produced by hard

classifying the original image. However, the method is satisfactory to evaluate the relative performance of the SRM algorithm in terms of the real image conditions due to rare availability of the sub-pixel reference data.

As a matter of fact, the input of PS algorithm (the proportional information) plays a decisive role in accuracy of the final super resolved map. In this research, the used LSMM involves several uncertainties related to end-member selection and linearity assumption of the model. So, exploring different soft classification methods through SRM procedure (e.g. neural networks, support vector machine, etc.) is suggested as an area for further investigation. Likewise, evaluation of other SRM algorithms and their influential parameters still needs more researches.

References

Atkinson PM (2004) Super-resolution land cover classification using the two-point histogram, In: GeENV IV: geostatistics for environmental applications, pp 15–28

Atkinson PM (2005) Super-resolution target mapping from soft classified remotely sensed imagery. Photogramm Eng Remote Sens 71(7):839–846

Atkinson PM (2009) Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study. Int J Remote Sens 30(20):5293–5308

Atkinson PM, Cutler MEJ, Lewis H (1997) Mapping sub-pixel proportional land cover with AVHRR imagery. Int J Remote Sens 18:917–935

Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203

Cai S, Du Q, Moorhead R (2007) Hyperspectral imagery visualization using double layers. IEEE Trans Geosci Remote Sens 45(10):3028–3036

Cai S, Moorhead R, Du Q (2011) An evaluation of visualization techniques for remotely sensed hyperspectral imagery. In: Prasad S, Bruce LM, Chanussot J (eds) Optical remote sensing: advances in signal processing and exploitation techniques. Springer, Berlin

Foody GM, Cox DP (1994) Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions. Int J Remote Sens 15:619–631

Foody GM, Muslim AM, Atkinson PM (2005) Super-resolution mapping of the waterline from remotely sensed data. Int J Remote Sens 26(24):5381–5392

Garcia-Haro FJ, Gilabert MA, Meliá J (1996) Linear spectral mixture modelling to estimate vegetation amount from optical spectral data. Int J Remote Sens 17:3373–3400

Holben BN, Shimabukuru YE (1993) Linear mixing model applied to coarse spatial resolution data from multispectral satellite sensor. Int J Remote Sens 14:2231–2240

Jacobson NP, Gupta MR, Cole JB (2007) Linear fusion of image sets for display. IEEE Trans Geosci Remote Sens 45(10):3277–3288

Kasetkasem T, Arora MK, Varshney PK (2005) Super-Resolution land cover mapping using a Markov random field based approach. Remote Sens Environ 96(3–4):302–314

Marcal A (2005) Automatic color indexing of hierarchically structured classified images. In: Proceedings of IEEE geoscience and remote sensing symposium, vol 7, pp 4976–4979

Mertens KC, Verbeke LPC, Ducheyne EI, De Wulf RR (2003) Using genetic algorithms in sub-pixel mapping. Int J Remote Sens 24:4241–4247

Niroumand JM, Safdarinezhad AR, Sahebi MR, Mokhtarzade M (2012) A novel approach to super resolution mapping of multispectral imagery based on pixel swapping technique. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci I-7:159–164. doi:10.5194/isprsannals-I-7-159-2012

Paola JD, Schowengerdt RD (1995) Review article: a review and analysis of back propagation neural networks for classification of remotely sensed multispectral imagery. Int J Remote Sens 16:3033–3058

Richards JA, Jia X (2006) Remote sensing digital image analysis, 4th edn. Springer, Heidelberg

Tatem AJ, Lewis HG, Atkinson PM, Nixon MS (2002) Super-resolution land cover mapping from remotely sensed imagery using a Hopfield neural network. In: Foody GM, Atkinson PM (eds) Uncertainty in remote sensing and GIS. London, Wiley

Thomas LL, Benning VM, Ching NP (1987) Classification of remotely sensed images. Adam-Hilger, Bristol

Thornton MW, Atkinson PM, Holland DA (2006) Super-resolution mapping of rural land cover features from fine spatial resolution satellite sensor imagery. Int J Remote Sens 27:473–491

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(2):234–240

Villa A, Chanussot J, Benediktsson JA, Jutten Ch (2011) Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution. IEEE J Sel Top Signal Process 5(3):521–533

Wessels R, Buchheit M, Espesset A (2002) The development of a high performance, high volume distributed hyperspectral processor and display system. In: Proceedings of IEEE geoscience and remote sensing symposium, vol 4, pp 2519–2521

A Digital Watermark Algorithm for Tile Map Stored by Indexing Mechanism

Na Ren, Chang-qing Zhu, Shu-jing Ren and Yi-shu Zhu

Abstract With the fast development of various applications for map service, the construction and development of the site of Google map have attracted a growing interest from the scientific and industrial communities recently. However, while dealing with tile map in web service publication, a crucial issue of copyright protection arises. The purpose of this chapter is to propose a novel mapping mechanism based watermarking algorithm for tile map, in such a way that the invisible and robust watermark information could be embedded effectively. Thus, the confidential and secure communication of tile map is obtained in the network. The characteristics and requirements of tile map stored by the indexing mechanism are analyzed firstly, which could help the application of the watermark embedding. Then, the watermark generated by m sequence is embedded in the blue channel by using the mapping mechanism. Finally, the experimental results are given to confirm that the proposed algorithm is robust to shearing, splicing, rotating, and additive noise.

N. Ren (✉) · C. Zhu · Y. Zhu
Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing 210046, Jiangsu Province, China
e-mail: renna1026@163.com

S. Ren
Xi'an Division of Surveying and Mapping, Xi'an 710054, Shaanxi Province, China

1 Introduction

With the development of network techniques, the geographic data is more widely used. Many websites, such as Google and Baidu map have provided the online electronic map services. Particularly, the geomatics industry is greatly pushed forward by the popularity of Google Earth, which is the representative software for publishing geographic data. China is also making great efforts to develop the web geographic information service, which indicates that the surveying and geomatics industry has been going towards wider public geographic information service in China. It will bring opportunities for broadening service space and promoting industrial development.

In the network environment, the geographic data is usually stored in the form of tiles at service terminals. This mode causes a severe problem that the data is easy access to downloading. Thus, the resulted copyright problems like illegally downloading, unauthorized using or transmitting via the Internet and unlawful earning become more and more serious. It greatly damages the copyright of data owners and harasses geomatics industry from healthy development. According to the previous research works, the consumers are more and more concerned about the security of privacy and copyright (Jahangir and Begum 2008; Alanazi et al. 2010; Alam and Siraj 2010). Hence, it has become an urgent problem to find a solution to protect the copyright of tile map and deal with conflicts between sharing tile map and keeping data security in the network environment. The digital watermark technique is considered as an efficient approach to solve this problem. It could provide reliable methods for data copyright protection, tracking data and checking the origin of illegal data (Kbaier and Belhadj 2006; Zhu et al. 2010).

This chapter is concerned with the requirements imposed by the tile map on watermark techniques, in the case of copyright protection. The improved algorithm of watermarking techniques is proposed based on literature (Ren et al. 2011), and the possible attacks of a watermarking algorithm are evaluated in terms of specific issues of tile map. The rest of the paper is organized as follows. In Sect. 2, the characteristics and requirements of tile map are analyzed. In Sect. 3, the proposed watermark algorithm is described. Section 4 is devoted to the presentation of experimental results. Finally, some conclusions are drawn in Sect. 5.

2 Characteristic Analysis of Tile Map

When a map on a certain scale is sliced to several rows and columns of square grid pictures according to the appointed size (usually 128×128 or 256×256 pixels), each square raster picture is called as a piece of tile (Yin and Sun 2010). Similar to raster maps, the tile map is stored by index mechanism, which is an approach based on pallet, and could save the space for storing masses of data. For indexed images, the colors' index values are stored in image file data. Because of the large

differences among these colors and the small amount of color data, the correlation of color data in spatial position is almost lost.

The data stored by indexing mechanism for the digital watermarking is based on an image color table. Compared to the grayscale or true color images, the digital watermark for tile map offers the typical characteristics:

1. The specific colors can be chosen and displayed conveniently according to the corresponding index values in the image color table.
2. Each piece of tile map contains less information than a raster map, and the amount of colors used in indexing is very limited, which makes it really difficult to hide information into data stored by indexing mechanism.
3. The sequence of values in the indexing table does not affect the display of tile map, so the correlation in spatial position doesn't exist.
4. Slight changes of values in the indexing table do not affect the display or analysis of tile map.

Currently, most watermark algorithms for digital raster map are based on grayscale or true color maps. On the other hand, the research works on watermark algorithm stored by indexing mechanism are very seldom (Wang et al. 2006; Fu et al. 2011; Zhu et al. 2009; Hu et al. 2005). For watermark algorithm of indexed images, the solution available is to transform the indexed data into the true color images at first. Then the true color images are handled with watermark algorithm (Wang et al. 2002). Obviously, the method of embedding watermark information in frequency domain is not appropriate for it. By improving the work in Ren et al. (2011), this chapter proposes and implements a novel watermarking algorithm in space domain for the tile map copyright protection.

3 Digital Watermark Algorithm for Tile Map

3.1 Watermark Information Generating

The watermark technique for tile map copyright protection is required to involve the copyright information of tile map. So, the copyright information should be mapped to the meaningless watermark information. The theory of m sequence is mature, and the output status is directly determined by the original status and the feedback logic. It possesses the typical features of uniqueness, proportionality and good autocorrelation, which are very suitable for the watermark information.

So, the two-valued bits are generated by the m sequence for embedding the meaningless watermark information. Then, the copyright information is mapped to those two-valued bits one by one, and the correlation is thus established. When detecting the watermark, the correlation coefficient between the watermark information and the two-valued bits generated by m sequence is figured out. If the correlation coefficient is bigger than the given threshold, the copyright information

of tile map is extracted according to the relationship between the watermark seed
and the copyright information.

3.2 Watermarking Algorithm Based on Mapping Mechanism for Tile Map

In literature (Ren et al. 2011), a digital watermark algorithm is put forward based
on the mapping mechanism for remote sensing images. The algorithm determines
the embedding location of watermark information by building the mapping
functions between the image data and the watermark information. The mapping
variables are expanded during the process. However, the values in color chart of
tile map stored by indexing mechanism do not mean specific colors' lightness
values, but addresses of color chart entrance instead. Thus, the values in color
chart will no longer have anything to do with correlation in spatial position.
Therefore, the method in Ren et al. (2011) can not be applied to design a mapping
function for the index mechanism based watermarks.

3.2.1 Watermark Embedding

Considering the correlation among three color values and the fact that the three
colors could describe one single dot, the two variables of mapping function are
designed to be two of the three color values RGB (R: red; G: green; B: blue). After
transforming the RGB color model into the $NTSC$ standard, the image data will be
made up of three parts: brightness (Y), tone (I) and saturation (Q). Y is the
brightness information. I and Q stand for the colorful information. The transform
formula is:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{1}$$

The scope of the three channels is [0,255], so, the impact of the Y can be
calculated as follows:

$$\begin{cases} 0.299 * R \in [0, 76.25] \\ 0.587 * G \in [0, 149.69] \\ 0.114 * B \in [0, 29.07] \end{cases}. \tag{2}$$

From the formula above, it can be observed that channel G in RGB affects the
brightness information in the $NTSC$ standard most, while channel B has the least
impact. So, it can be inferred that modification of channel B will only make the
brightness information Y change for 29.07 at most. Furthermore, channel R and

G are designed as the mapping variables, and the corresponding mapping function $f(R,G) = I_W$ for the watermark information is designed.

The detailed model of watermark embedding for tile map based on indexing mechanism is proposed as follows. It is assumed that the data set stored by indexing mechanism of channel R is $R = \{R_{ij}, 0 \leq i < M, 0 \leq j < M\}$, and that of channel G is $G = \{G_{ij}, 0 \leq i < M, 0 \leq j < M\}$. Their lengths are both $M \times M$. The watermark information sequence is $W = \{w_i, 0 \leq i < L\}$ with length L. After embedding watermark information into data set of channel B, the data set is turned to be B^w. The embedding rule is given as:

$$B^w = B \oplus W = \begin{cases} B_{ij}^w = b_{ij} \oplus \left(f(R_{ij}, G_{ij}) * \delta * w_{f(R_{ij}, G_{ij})} \right), & g(b_{ij}) = TRUE \\ B_{ij}^w = b_{ij}, & g(b_{ij}) = FALSE \end{cases} \quad (3)$$

where $\oplus$ is the watermark embedding rule, and the function $f(R, G)$ is the mapping relationship between the tile map and the watermark information. $W_{f(R,G)}$ is the corresponding watermark information extracted by the mapping function. δ is the embedding strength of tile map b_{ij}, and $g(b_{ij})$ indicates whether tile map b_{ij} can be embedded with watermark. *TRUE* means the watermark can be embedded, while *FALSE* means the watermark embedding can not be done.

3.2.2 Watermark Detecting

Watermark detecting is the inverse process of embedding. The watermark detecting can be performed without knowledge of the original watermark information. Instead, the random number generated is needed. Similar to the calculation in literature (Ren et al. 2011), it is based on judging the correlation coefficients to determine if a watermark is embedded or not. The details on the optical detection scheme employed are listed in the earlier work by literature (Ren et al. 2011), so the detailed description is omitted here.

4 Experiments and Analysis

In this section some experimental results are given to evaluate the performance of the proposed algorithm. The message "COPYRIGHT" is converted into the two-bit values generated by m sequence. The resulted bits are embedded in the blue channel. The tile map may be subject to two kinds of attacks, namely the intentional attack and the unintentional attack, such as splicing, shearing, rotating, noise addition, filtering, removal attempts and desynchronization attacks. In this chapter, we consider splicing, shearing, rotating, and noise addition, since these attacks appear in practical application most frequently.
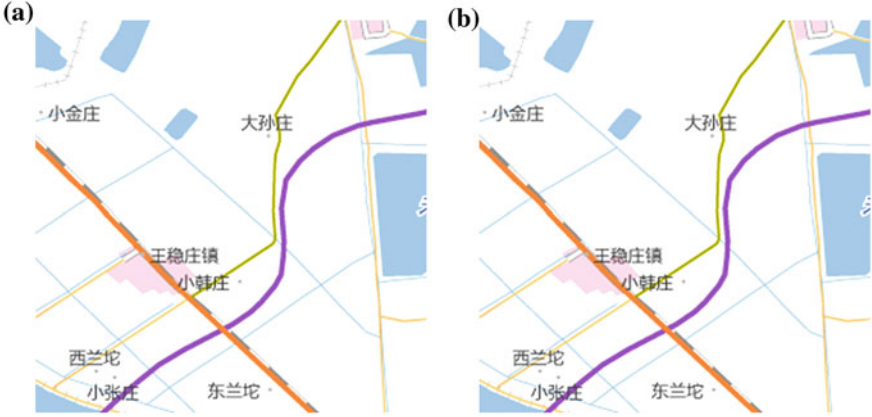
Fig. 1 Example 1. **a** Original tile map. **b** Watermarked tile map

This chapter uses Peak Signal to Noise Ratio (PSNR) as the distortion measure. The PSNR is expressed mathematically in the form as follows.

$$PSNR = \frac{XY \max I^2(x, y)}{\sum_{x,y} \left[I(x, y) - \tilde{I}(x, y) \right]^2} \tag{4}$$

where $I(x, y)$ represents a pixel value whose coordinates are (x, y) in the original tile map, and $\tilde{I}(x, y)$ represents a pixel value whose coordinates are (x, y) in the watermarked tile map. The numbers of rows and columns in the pixel matrix are denoted by X and Y.

1. Invisibility
   Figures 1, 2 show two examples for comparison, where Figs. 1a, 2a show the original tile maps without watermark, while Fig. 2a, b shows the watermarked tile maps. The PSNR value between the watermarked and original tile map in Fig. 1 is 41.685 dB, while that in Fig. 2 is 41.369 dB, respectively. Furthermore, it is very difficult to tell the difference between the original and watermarked tile map in the subjective visual. So, the invisibility of this algorithm is verified well.

2. Robustness
   For verifying the robustness of the proposed algorithm, the watermarked tile maps are sheared, rotated, spliced and noise addition. In the experiments, the tile maps are sheared at any place and in any size, and they are also spliced in a random way. Not only the two tiles with the same watermark are spliced effectively, but also the tile embedded with watermark and that without watermark could be spliced successfully. The tile maps are also verified to maintain good quality when they are rotated in an arbitrary angle and added noises. The experimental results show that the watermark information can still be detected correctly even after these attacks. So the algorithm proposed in this chapter is verified to be robust.
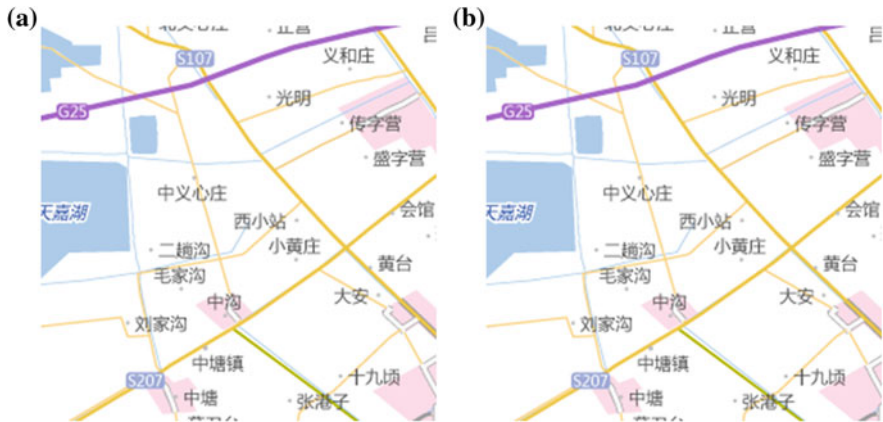
Fig. 2 Example 2. **a** Original tile map. **b** Watermarked tile map

5 Conclusions

A novel modified mapping mechanism based digital watermarking algorithm is proposed for tile map in this chapter. The characteristics of tile map stored by indexing mechanism and their requirements of watermark algorithm are analyzed at first. Then, the watermark algorithm is designed based on the improved mapping mechanism. The experimental results are given to verify that the proposed watermark algorithm offers the advantages of good invisibility and robustness. It is able to meet the requirements of tile map copyright protection in current web environments, and will have good prospect of application.

References

Alam GM, Siraj S (2010) Information and communication technology for participatory based decision-making-E-management for administrative efficiency in higher education. Int J Phys Sci 5(4):383–392

Alanazi HO, Jalab HA et al (2010) Securing electronic medical records transmissions over unsecured communications: an overview for better medical governance. J Med Plants Res 4(19):2059–2074

Fu H, Zhu C et al (2011) Multipurpose watermarking algorithm for digital raster map based on wavelet. Acta Geodaetica et Cartographica Sinica Transform 40(3):397–400

Hu Y, Wu H, Zhang H et al (2005) A digital watermarking scheme in huge index image. J Beijing Univ Posts Telecommun 28(1):26–29

Jahangir N, Begum N (2008) The role of perceived usefulness, perceived ease of use, security and privacy, and customer attitude to engender customer adaptation in the context of electronic banking. Afri J Bus Manage 2(1):032–040

Kbaier I, Belhadj Z (2006) A novel content preserving watermarking scheme for multispectral images. Inf Commun Technol 16(2):243–247

Ren N, Zhu C, Wang Z (2011) Blind watermarking algorithm based on mapping mechanism for remote sensing image. Acta Geodaetica et Cartographica Sinica 40(5):623–627

Wang L, Dong X, Mei Z et al (2002) Research and realization of digital watermarking for image and wave. J Northeast Univ (Nat Sci) 23(3):217–220

Wang X, Zhu X, Bao H (2006) Complementary watermarking algorithm for digital grid map. J Zhejiang Univ (Eng Sci) 40(6):1056–1059

Yin F, Sun L (2010) The research of map publishing platform development based on the tile pyramid technology. Geomat Spatial Inf Technol 33(5):16–17

Zhu C, Fu H, Yang C (2009) Watermarking algorithm for digital grid map based on integer wavelet transformation. Geomat Inf Sci Wuhan Univ 34(5):619–621

Zhu C, Yang C, Ren N (2010) Application of digital watermarking to geospatial data security. Bull Surv Mapp 10:1–3

Part III
Generalisation and Update Propagation

Towards Cartographic Constraint Formalization for Quality Evaluation

Xiang Zhang, Tinghua Ai, Jantien Stoter and Jingzhong Li

Abstract This paper presents a first-order representation to formalize carto-graphic constraints for automated quality evaluation of multi-scale data. Formal-izing constraints for cartographic applications is a challenging task. It requires precise definition of entities, spatial and semantic relationships for individuals, groups and classes of objects, and their (intra-/inter-scale) relationships. Also constraints defining the visual presentation of the same entities can be different depending on the scale and context. This paper categorizes and formalizes dif-ferent types of information needed for the quality evaluation, based on which cartographic constraints are formalized. The formalism is demonstrated by applying it to group features such as networks and alignments, and finally to constraints of different levels of complexity. We show the potential of the pro-posed formalism and discuss possibilities for further development.

Keywords Formal map specifications · Multi-scale modeling · Automated evaluation of generalized maps

1 Introduction

Spatial data is being created and maintained at various scales to meet the increasing demand of geo-information. The increase of data volume and update speed has led to heterogeneous representations of geographic areas. This inevitably causes inconsistencies between multiple representations. In generalization domain,

X. Zhang (✉) · T. Ai · J. Li
School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
e-mail: xiang.zhang@whu.edu.cn

J. Stoter
GISt, OTB, Delft University of Technology, Delft, the Netherlands

cartographic constraints are widely used to express user expectations concerning the quality of maps (including data and visual qualities) and hence are used to control and evaluate the quality of generalization.

Various models of constraints were proposed for map generalization and quality evaluation (e.g. Weibel and Dutton 1998; Burghardt et al. 2007). This research is a continuation of our previous work (Zhang et al. 2008). Also Touya et al. (2010) proposed a model of constraints for generalization which shares similar aspects. Note that, constraints for the evaluation differ (slightly) from those for generalization. First, constraint priority, the sequence in which constraints are evaluated, is not relevant for the evaluation of generalized results (Zhang 2012). Second, preferred action(s) are needed for modeling constraints for generalization (Burghardt et al. 2007; Stoter et al. 2009) but not for the automated evaluation. Third, corresponding relationships between objects of different scales must be explicitly expressed for the evaluation.

Besides, constraints were previously formulated in a more or less informal way, which brings about confusions in their interpretation, even for human experts (Burghardt et al. 2007; Zhang et al. 2008). Rather than proposing a new model of constraints, this paper presents a formalism that refines the existing models and provides a foundation for the machine-based interpretation of constraints. The ultimate goal is that, by interpreting the constraints, the machine 'knows' how to evaluate it, which high-level concepts, relations and contexts to detect, which algorithms/operations to use, etc. This formalism should also be useful for automated generalization. In the following, we firstly present an object-oriented model for relevant spatial entities and relationships (Sect. 2), where high-level concepts and relationships are exemplified. Section 3 formalizes cartographic constraints into first-order logic expressions. Section 4 discusses the expressive power of the formalism, implementation issues and possibilities for automated quality evaluation. More details and design principles are discussed in the doctoral thesis of Zhang (2012).

2 Spatial Entities and Relationships

This section formalizes some basic concepts pertinent to the formalization of cartographic constraints. First, we distinguish between three categories of entities: individuals (objects), universals (classes), and collections (groups). This is widely received in automated generalization (Ruas 2000; Stoter et al. 2009) and is also in line with findings in ontology research (Bittner et al. 2004; Mark et al. 2001). Then we give formal definitions of spatial entities in a multi-scale setting (Sect. 2.1) and describe the relationships between different entities (Sect. 2.2). Several examples of the formalism are shown in Sect. 2.3.

2.1 Entities, Objects and Groups

An *entity* can be either an object (feature) or a group of objects (Fig. 1). It is formally given by a 2-tuple:

$$\text{Entity} = \langle ID, Granu \rangle \tag{1}$$

where:

- *ID* is the identifier of the entity (object or group) being referred to;
- *Granu* ∈ {*micro, meso, macro*} indicates its granularity.

Here granularity specifies the degree to which the objects are grouped. For example, a building object is a micro entity; objects grouped by a partition or into an alignment belong to a meso entity; a feature class of objects (e.g. all buildings in a dataset) is a macro entity.

Object and group of objects are two fundamental concepts derived from entity. An *object* is an individual which can be given by a triple:

$$\text{O} = \langle ID, CL, Geo \rangle \tag{2}$$

where:

- *CL* is the feature class, e.g. *Building* or *Road*;
- *Geo* is the geometry of the object, which is composed of: (i) geometry types *Geo type* ∈ {*point, linear, areal*} and (ii) the spatial representation (coordinate pairs) *Geo. rep* = $\{p_1, ..., p_n\}$.

We identify groups as explicit entities with identifiers, because groups have structural properties (e.g. density, spatial distribution and other statistical properties) which should be preserved during the generalization. A *group* is a collection of objects, and can be given by a 4-tuple:

$$\text{Gr} = \langle ID, OC, R, Geo \rangle \tag{3}$$

where:

- $OC = \{O_1, ..., O_n\}, n \geq 0$ is a collection that forms the group;

Fig. 1 UML diagram of entity, object, group and their relations (Zhang 2012)
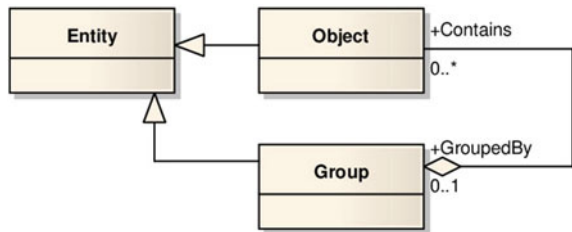
- *R* is the relationship that holds among the objects in the collection;
- *Geo* is the representative geometry of the group.

This definition is flexible enough to cover various kinds of group entities. First, a group can be a collection where some relationship holds for its members (e.g. alignment), or between the members and some other objects (e.g. neighborhood and partition relationship). Second, the above definition also allows groups of objects from different feature classes that create 'higher-level' phenomena (e.g. a 'housing aggregate' composed of a house, a garden and an outbuilding). The group entity in the second sense is similar to the concept of composite object described in (van Smaalen 2003).

Finally, a *feature class* only consists of meta information and is given by FeaCls = $\langle Cls, Geotype \rangle$, where $Cls \in \{Building, Road,...\}$ and $Geotype \in \{point, linear, areal\}$. '*Geotype*' is different from '*Geo*' in the Object definition (*Def.* 2) which has type and geometry properties.

The semantic relationship describing that an object is an instance of a feature class can be defined as:

$$
\begin{aligned}
InstanceOf(O_i, FeaCls_i) \\
\equiv (O_i.CL = FeaCls_i.Cls) \\
\wedge (O_i.Geo.type = FeaCls_i.Geotype).
\end{aligned}
\tag{4}
$$

Now, we can express all instances of a feature class C with a specialized group:

$$
Gr_C = \langle ID, \{O_i | (\forall O_i)(InstancOf(O_i, C))\}, \emptyset, Geo \rangle
\tag{5}
$$

Here the object collection (Gr$_C$. *OC*) is given by a set-builder notation, meaning that for all objects in the domain of discourse, anything that is an instance of class C is an element of Gr$_C$. *OC*. Hence Gr_C is actually the extension of class *C*. In a specific application, Gr_C can be used to indicate all the instances of a class that exist in a particular dataset.

Note in *Def.* 5 that we set Gr$_C$. $R = \emptyset$ since the relation is already defined in the set-builder notation of Gr$_C$. *OC*. For meso groups, a non-empty relationship *R* (see *Def.* 3) should be explicitly specified to explain how the group members are related (see examples in Sect. 2.3).

We also use the following predicate syntax to denote that O_i is an instance of some class, where <class_name> is a placeholder (e.g. *Built_up_area*(O_i) means O_i is a built-up parcel):

$$
<class\_name>(O_i) \equiv InstanceOf(O_i, <class\_name>)
\tag{6}
$$

2.2 Spatial Relationships Between Entities

This section focuses on a set of relationships between individuals (i.e. spatial entities) that is relevant to the proposed evaluation. We further distinguish between the relationships and properties within the same scale (intra-scale relationships) and those of different scales (inter-scale relationships).

Intra-scale relationships between individuals in the geospatial domain **D** can be characterized in terms of qualitative and quantitative relationships. The relationships in topological space are usually qualitative, whereas the ones in metric space are usually quantitative in nature.

We choose to formalize spatial relationships using functions instead of predicates or mathematic relations. Because predicates or mathematic relations are mostly qualitative, so they cannot describe the continuous change of spatial properties over scale transitions. Besides, although quantitative relationships can be qualified, the qualification process can be cognitively plausible and context dependent (Worboys 2001). The functional specification of spatial relationships is given below:

$$
\begin{aligned}
f \quad &: \quad \mathbb{D}^n \to \mathbb{R}^n &&\text{(Quantitative)} \\
&: \quad \mathbb{D}^n \to \{\text{true}, \text{false}\} &&\text{(Qualitative)} \qquad (7) \\
(x_1, &\ldots, x_n) \mapsto f(x_1, \ldots, x_n) &&
\end{aligned}
$$

where $\mathbf{D}^n$ is the domain of function f which takes $x_1,\ldots,x_n$ as its input (x_i is a spatial entity); $\Re^n$ is the output space of f. Although $\Re$, or more precisely $\Re^+$, is in most cases sufficient, $\Re^n$ is specified to allow for a larger degree of flexibility. Ai et al. (2008) for instance presented a Fourier transform based shape descriptor, which yields a series of numbers to characterize shapes. This case can be formalized in a function: $FourierShape(x) = (c_1, \ldots, c_n)$.

This functional view provides a unified approach to the formalization of qualitative and quantitative relationships. Qualitative relationships in particular can be seen as functions with Boolean output. Examples of relationships and properties in a functional form are $Size(x_1) = 4\,\text{mm}^2$, $Distance(x_1, x_2) = 3\,\text{mm}$ and $Contain(x_2, x_3) = true$.

Contextual relationships are special intra-scale relationships. According to Mustière and Moulin (2002), contextual relationships can be divided into: (i) being part of a significant group (e.g. alignments), (ii) being located in a particular area, and (iii) being in relation to the 'same-level' entities.

The contextual relationship here only means 'being located in a particular area'. Being part of a significant group can be expressed by the *member-of* relationship between a group and its members (see *Def.* 3). The third type of relationships is formalized as normal intra-scale relationships (*Def.* 7). The concept of 'same-level' is defined by our division of entities into individuals, collections and universals. The contextual relationship that an object is located in a particular context can be defined as follows:

Fig. 2 Inter-scale relationships between objects, groups and classes (Zhang 2012)
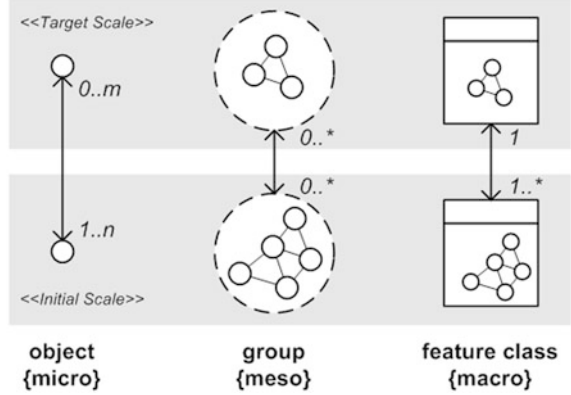
$$InContext(x) \equiv (\exists c)(Context(c) \wedge Contain(c, x)). \tag{8}$$

Our definition of spatial context requires that $c.CL = <Context>$ and $c.Geo.-type = areal$. For example, in describing a building b_1 located in urban areas, we write $InUrban(b_1)$, meaning that there exists an instance c_1 such that both $Urban(c_1)$ (see *Def.* 6) and $Contain(c_1, b_1)$ hold. Note that we omit the context instance c from the notation $InContext(x)$, because we only concern what type of context x is located in but not which context instance. Uncertainty issues should be handled in measuring contextual relationships.

Inter-scale relationships. Spatial entities of different scales that represent the same real-world phenomena are related (linked) by inter-scale relationships (Bobzien et al. 2008). The inter-scale relationships between entities of different granularities are illustrated in Fig. 2.

The inter-scale relationships have varying multiplicities depending on the granularity of involved entities. For individual objects, 1-to-1 or 1-to-0 is possible if object was kept or removed by generalization; many-to-many (n-to-1 or n-to-m) relationships are caused by aggregation or typification. Alternatively, n-to-m relationship between micro-objects can be modeled by 1-to-1 relationship between meso-groups. For feature class level entities, 1-to-1 relationship is the common case, though sometimes several classes may be merged into one class (n-to-1).

Here, we decompose the many-to-many relationship into a set of 1-to-1 relationships. The same strategy was discussed in Bobzien et al. (2008). We call it *corresponding* relationship and define it as a binary relationship:

$$Corr\left(E_i^{Sp}, E_j^{Sq}\right) \quad : \mathbb{D}^2 \rightarrow \{\text{true}, \text{false}\} \qquad (\text{Certain})$$
$$: \mathbb{D}^2 \rightarrow [0, 1] \qquad\qquad (\text{Uncertain}) \tag{9}$$

where Sp and Sq indicate the scales of the respective entities; p and q are scale variables. The corresponding relationships created by generalization are regarded as certain links; whereas those created by data matching are uncertain links (with matching probabilities).

2.3 Meso-Groups as Examples

In this section, examples are given on meso-level groups, namely networks and alignments, using the formalism described in Sects. 2.1 and 2.2.

Networks are characterized by elementary properties such as connectivity and collective ones such as spatial distributions and patterns (e.g. dendritic pattern). A network feature is given by a group entity with a connectivity relationship:

$$Gr_{net} = \langle ID, OC = \{o_1, \ldots, o_n\}, Connectivity(OC), Geo \rangle \qquad (10)$$

where $o_i.Geo.type = linear$ is required; Geo is the representative geometry of the network (e.g. catchments of rivers). Usually, $Gr_{net}.Geo.type = areal$.

By introducing the concept of $path$, p, we can define the relationship connectivity as:

$$Connectivity\Big(OC\Big) \equiv \Big(\forall o_i, o_j\Big)\Big(\Big(o_i, o_j \in OC\Big)$$
$$\Leftrightarrow Touch\Big(o_i, o_j\Big)$$
$$\vee \Big(\exists p\Big)\Big(\Big(p \subseteq OC \wedge Touch\Big(o_i, p\Big) \wedge Touch\Big(o_j, p\Big)\Big)\Big)$$

where $\Leftrightarrow$ is a bidirectional logical implication; $p = \{o_1, \ldots, o_m\}$, $m < n$ is a sequence of connected spatial features such that $(\forall o_i, o_{i+1})((o_i, o_{i+1} \in p) \Rightarrow Touch(o_i, o_{i+1}))$. This definition also allows for connected polygons.

More generally, we can define a binary relationship that hold between connected objects:

$$Connected\big(o_i, o_j\big) \equiv (\exists p)\big(Touch(o_i, p) \wedge Touch\big(o_j, p\big) \vee Touch\big(o_i, o_j\big)\big). \quad (11)$$

Alignments are linear arrangements of individuals (e.g. buildings, ponds, islands). Such linear patterns should be kept in generalization (Christophe and Ruas 2002; Stoter et al. 2009). An alignment is given by a group entity with a relationship called *Aligned*:

$$Gr_{align} = \langle ID, \{o_1, \ldots, o_n\}, Aligned(o_1, \ldots, o_n), Geo \rangle \qquad (12)$$

where Geo is usually represented by the skeleton of the alignment. This skeleton can be used to indicate the virtual (curve) line to which the group members align, so we require that $Geo.type = linear$. The spatial relationship, *Aligned*, that glues the group members together is based on a general notion of consistency and can be defined recursively:

$$Aligned(o_1, o_2) = Consistent(\{o_1\}, o_2)$$
$$Aligned(o_1, o_2, o_3) = Consistent(\{\{o_1\}, o_2\}, o_3)$$
$$\cdots = \cdots$$
$$Aligned(o_1, o_2, \ldots, o_n) = Consistent(\{\ldots\{\{o_1\}, o_2\}, \ldots\}, o_n)$$

For the semantics of *Consistent* one is referred to Zhang et al. (2013). Major properties of alignments include spatial distribution and orientation. Alignments or clusters are regular spatial distributions, so we use homogeneity (Christophe and Ruas 2002; Zhang et al. 2013) to quantify the degree of regularity of alignments or clusters. It can be formally expressed as $Homo(align_i) : \mathbf{D} \to [0,1]$. Similarly, $Orientation(align_i) : \mathbf{D} \to \Re^+$ expresses the orientation of $align_i$. Parallel relationship between alignments, $Parallel(align_i, align_j) : \mathbf{D}^2 \to [0,1]$, is also useful if two alignments (especially curvilinear ones) are to be compared in terms of orientation.

3 Constraints as First-Order Expressions

In expressing cartographic constraints, several side notes should be considered. First, the domain of discourse $\mathbf{D}$ is subdivided into sub-domains $\mathbf{D}^{S1}, \ldots, \mathbf{D}^{Sn}$, distinguishing between entities at different scales. So o^{S1} and Gr^{S2} are an object and a group of different scales. Second, aliases are frequently used to simplify the expressions (e.g. $align_1$ is a Gr_{align}). Third, the conditions to be respected by constraints are formalized as conditional expressions, e.g., $Size(o_1) \geq \mu$ where μ is a user-defined parameter. In the following, we formalize a subset of constraints drawn from NMA specifications (Stoter et al. 2009) using a first-order language.

To begin with, a minimum dimension constraint (C1) saying that area of any target polygon should be larger than a certain threshold is formalized as:

$$(C1)\left(\forall o_i^T\right)\left(\left(o_i^T.Geo.type = areal\right) \Rightarrow Size\left(o_i^T\right) \geq \mu\right).$$

Note that, o_i^T, with T denoting target scale, is used to simplify the expression $(o_i)(o_i \in \mathbf{D}^T)$ in this context. Likewise, o_i^I denotes an object at the initial scale. The second constraint (C2) is a requirement saying that important initial buildings should be kept:

$$(C2)\left(\forall b_i^I\right)\left(ImportantBuilding\left(b_i^I\right) \Rightarrow \left(\exists b_i^T\right)\left(Corr\left(b_i^I, b_i^T\right)\right)\right)$$

where b_i is an alias of an instance of building class in this case; the type predicate (*Def.* 6) restricts the range of the quantifier to important buildings. C2 explicitly states that for every important building in the initial data, there exists at least one building in the target data that correspond to it.

The third constraint (C3) is a preservation constraint with added contexts (scopes). It requires that target shape should remain concave if the initial polygons are of high concavity:

$$(C3)\left(\forall a_i^I\right)\left(\left(Concavity\left(a_i^I\right) \geq \mu\%\right) \wedge \left(\exists a_i^T\right)\left(Corr\left(a_i^I, a_i^T\right)\right) \right.$$
$$\left. \Rightarrow Similar\left(Concavity\left(a_i^I\right), Concavity\left(a_i^I\right)\right)\right).$$

Note that, the degree of concavity *Concavity* should be measured at an implementation level, and 'high concavity' is application dependent (e.g. $\mu = 95\ \%$). We use a predicate *Similar* here since its concrete form is unknown at this moment. Note that the scope that limits the polygons in the initial data states that C3 only considers the initial objects (i) that are highly concave in shape and (ii) that were kept in the target data.

The next example shows a constraint requiring that proximate and roughly parallel features (line–line or polygon–polygon) should become topologically adjacent (C4):

$$(C4)\left(\forall o_i^I, o_j^I\right)\left(\left(Distance\left(o_i^I, o_j^I\right) < \mu_1\right) \wedge \left(Parallel\left(o_i^I, o_j^I\right) \geq \mu_2\right)\right.$$
$$\wedge \left(\exists o_i^T, o_j^T\right)\left(Corr\left(o_i^I, o_i^T\right) \wedge Corr\left(o_j^I, o_j^T\right)\right.$$
$$\left.\left.\Rightarrow Adjacent\left(o_i^T, o_j^T\right)\right)\right)$$

where μ_1, distance, and μ_2, degree of parallelism, are to be specified in certain applications. Similar to C3, C4 only considers the initial object pairs where the two objects are close and parallel to each other and where both objects were kept in the target data. This is an example where the topological relationship between objects should change. Next, we exemplify that topological relationship (connectivity) should be preserved (C5) on top of the relationship Connected (*Def.* 11):

$$(C5)\left(\forall o_i^I, o_j^I\right)\left(Connected\left(o_i^I, o_j^I\right)\right.$$
$$\wedge \left(\exists o_i^T, o_j^T\right)\left(orr\left(o_i^I, o_i^T\right) \wedge Corr\left(o_j^I, o_j^T\right)\right.$$
$$\left.\left.\Rightarrow Connected\left(o_i^T, o_j^T\right)\right)\right)$$

where o can be either linear or areal features. C5 can be used to check whether the connectivity has been kept between any two objects or for network features. Consider the following variation of C5:

$$(C5')\left(\forall net_i^I\right)\left(\exists net_i^T\right) Corr\left(net_i^I, net_i^T\right)$$

where *net* is an alias of Gr_{net}. One may argue that the existence of a corresponding network in the target data automatically ensures the connectivity of the initial network. This is however not true as is shown in Fig. 3. In the illustration, the initial network was kept after generalization, but some river branches such as l_2 become disconnected from the initial network. Such a violation to connectivity can be identified by C5 but not C5'.

Finally, we formalize some constraints on alignments (al_i). The first one (C6) requires that alignment should be kept; the second one (C7) requires that for the target alignment its orientation should be similar to the initial one; the third one (C8) requires that the spatial distribution (regularity) of alignment members should be preserved or even enhanced.

Fig. 3 Example of connectivity: initial network (*left*) and generalized network (*right*) (Zhang 2012)
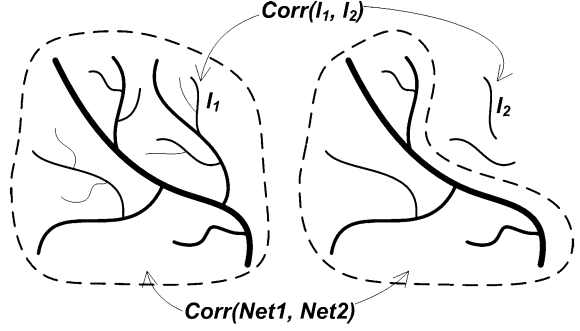
$$(C6)(\forall al_i)(al_i \subset D^I) \Rightarrow (\exists al_j)(al_j \subset D^T) \wedge Corr(al_i, al_j)$$

$$(C7)(\forall al_i^I)(\exists al_i^T)(Corr(al_i^I, al_i^T) \Rightarrow (|Orientation(al_i^I) - Orientation(al_i^T)| \leq \theta))$$

$$(C8)(\forall al_i^I)(\exists al_i^T)(Corr(al_i^I, al_i^T) \Rightarrow (Homo(al_i^T) \geq Homo(al_i^I))).$$

4 Discussion

4.1 Expressiveness of the Formalism

The formalism described is powerful enough to express constraints with different combinations of *entity class*, *scope modifier*, *constrained property* and *condition* to be respected. A detailed account of the syntactic structure of constraints estimates that in theory nearly 1,000,000 kinds of constraints can be expressed (Zhang 2012). The expressive power also comes from the ability to specify semantic, conditional or contextual scopes to limit the entities to be considered by a constraint. For example a constraint 'the road leading to a building in a peninsula should be kept' can be formalized as:

$$(C9)\left(\forall r_i^I b_j^I\right)\left(LeadTo\left(r_i^I, b_j^I\right) \wedge InPenisula\left(b_j^I\right)\right)$$
$$\Rightarrow \left(\exists r_i^T, b_j^T\right)\left(Corr(r_i^I, r_i^T) \wedge Corr\left(b_j^I, b_j^T\right)\right).$$

In this example, the range of quantification is specified through scope modifiers, i.e., the relationship *LeadTo* and the contextual relationship *InPenisula* (see *Def.* 8). This constraint states that for the initial roads and buildings that satisfy that buildings are located in peninsulas and roads lead to those buildings, there exist target objects that correspond to those initial roads and buildings (i.e. initial objects are kept).

4.2 Implementation and Use of the Formalism

This work provides a solid foundation for constraint formalization and is more geared towards a machine-based interpretation and reasoning of the constraint expressions, so that the meaning is clear to both human and computers. The proposed formalism is independent of implementation techniques or standards but can be translated into specific techniques. For example, the Object Constraint Language (OCL) (Warmer and Kleppe 2003) appears to have the potential to implement the presented formalism, since OCL roots in first-order predicate logic and it is able to express entities and relationships using UML constructs such as classes and methods. Research in geospatial domain has demonstrated the use of OCL and UML in modeling topological rules in spatial databases (Bejaoui et al. 2010) and specifications for MRDBs (Friis-Christensen et al. 2005; Stoter et al. 2011).

OCL is a more user-friendly language which requires much less mathematical understanding. To further bring down the barrier of use, we suggest designing GUIs according to the structure discussed in Zhang (2012) to enable a click-and-drag manner of constraint definition, which can then be translated into a machine readable format automatically.

4.3 Possibilities for Automated Quality Evaluation

It is interesting to see automated evaluation of generalized results in a logical framework, i.e., by interpreting the formal constraints and reasoning if they hold in data. This requires that all relevant entities, properties and relationships are stored in a database. Take constraint C9 for example, besides building and road instances, the objects that satisfy the *LeadTo*, *InPenisula* and *Corr* relationships (detected offline) should also be stored in relational tables. Because OCL constraints can be directly translated into SQL queries, C9 can be written in OCL in a way that buildings and roads that do not satisfy the constraints are returned.

However, due to the amount of information and the purpose-dependent nature of map generalization, not all information can be stored and some has to be enriched on-demand. In a fully formal approach, data enrichment operations could also be semantically annotated such that the logical unit can find the operations by semantic matching. In extreme cases high-level concepts such as peninsula can be defined by a semantic network or formal ontology, the machine can chain different operations to detect the concepts (e.g. Zhang et al. 2008). This vision is very challenging and requires quite a lot of research on the semantic aspect of geo-operations.

A disadvantage of using Boolean logic for automated evaluation is, however, that a constraint is always evaluated to be either violated or not. Soft evaluation (with violation degrees) is hence not possible. Given this shortcoming, the proposed formalism still provides a powerful way to precisely define cartographic

constraints. In addition, automated evaluation in a logical framework is still effective in identifying database errors and preventing undesired insertions into the database.

References

Ai T, Shuai Y, Li J (2008) The shape cognition and query supported by Fourier transform. In: Ruas A, Gold C, Cartwright W, Gartner G, Meng L, Peterson MP (eds) Headway in spatial data handling, LNGC, pp 39–54

Bejaoui L, Pinet F, Schneider M, Bédard Y (2010) OCL for formal modelling of topological constraints involving regions with broad boundaries. GeoInformatica 14(3):353–378

Bittner T, Donnelly M, Smith B (2004) Individuals, universals, collections: on the foundational relations of ontology. In: Proceedings of the 3rd international conference on formal ontology in information systems, pp 37–48

Bobzien M, Burghardt D, Petzold I, Neun M, Weibel R (2008) Multi-representation databases with explicitly modeled horizontal, vertical, and update relations. Cartogr Geogr Inf Sci 35(1):3–16

Burghardt D, Schmidt S, Stoter J (2007) Investigations on cartographic constraint formalisation. In: 10th ICA workshop of ICA commission on generalisation and multiple representation, Moscow

Christophe S, Ruas A (2002) Detecting building alignments for generalisation purposes. In: Richardson DE, Van Oosterom P (eds) Advances in spatial data handling. Springer, Heidelberg, pp 419–432

Friis-Christensen A, Jensen CS, Nytun JP, Skogan D (2005) A conceptual schema language for the management of multiple representations of geographic entities. Trans GIS 9(3):345–380

Mark DM, Skupin A, Smith B (2001) Features, objects, and other things: ontological distinctions in the geographic domain. In: Montello DR (ed) Spatial information theory: foundations of geographic information science. LNCS, pp 488–502

Mustière S, Moulin B (2002) What is spatial context in cartographic generalisation? In: Symposium on geospatial theory, processing and applications, volume 34-B4, pp 274–278

Ruas A (2000) The roles of meso objects for generalization. In: Proceedings of the 9th international symposium on spatial data handling, pp 3B50–3B63

Stoter J, Burghardt D, Duchêne C, Baella B, Bakker N, Blok C, Pla M, Regnauld N, Touya G, Schmid S (2009) Methodology for evaluating automated map generalization in commercial software. Comput Environ Urban Syst 33(5):311–324

Stoter J, Visser T, van Oosterom P, Quak W, Bakker N (2011) A semantic-rich multi-scale information model for topography. Int J Geogr Inf Sci 25(5):739–763

Touya G, Duchêne C, and Ruas A (2010) Collaborative generalisation: formalisation of generalisation knowledge to orchestrate different cartographic generalisation processes. In: Fabrikant S, Reichenbacher T, van Kreveld M, Schlieder C (eds) Geographic information science. LNCS, pp 264–278

van Smaalen J (2003) Automated aggregation of geographic objects: a new approach to the conceptual generalisation of geographic databases. PhD thesis, Wageningen University, Wageningen

Warmer J, Kleppe A (2003) The object constraint language: getting your models ready for MDA, 2nd edn. Addison-Wesley Longman Publishing Co., Inc., Boston

Weibel R, Dutton G (1998) Constraint-based automated map generalization. In: Proceedings 8th international symposium on spatial data handling, pp 214–224

Worboys MF (2001) Nearness relations in environmental space. Int J Geogr Inf Sci 15(7):633–651

Zhang X (2012) Automated evaluation of generalized topographic maps. PhD thesis, Twente University, Enschede

Zhang X, Stoter J, Ai T (2008) Formalization and automatic interpretation of map requirements. In: 11th ICA workshop of ICA commission on generalisation and multiple representation, Montpellier

Zhang X, Ai T, Stoter J, Kraak MJ, Molenaar M (2013) Building pattern recognition in topographic data: examples on collinear and curvilinear alignments. GeoInformatica 17(1):1–33

Preservation and Modification of Relations Between Thematic and Topographic Data Throughout Thematic Data Migration Process

Kusay Jaara, Cécile Duchêne and Anne Ruas

Abstract Nowadays, users often use topographic data to reference their own thematic data. When reference data are updated or if the user wants to replace the reference, the thematic data have to be processed in order to maintain data consistency. We call this processing thematic data migration. This paper proposes an updated version of a previously proposed thematic data migration process, which includes the case where the relations between thematic and topographic data have to be modified between the initial and the final state. A model to describe the relations and their modifications is proposed. A multi-criteria decision method is used to relocalise the thematic data on the topographic data guided by the described relations. The whole process is illustrated on a running example, on which obtained results are presented and discussed.

1 Introduction

Users create their own data using geographic data available from national mapping agencies or from volunteer geographic information. Users often use topographic data to reference their own thematic data. A user may be a professional operator or anyone who adds on his own geolocated information (e.g. user data from land survey). When reference data are updated or if the user wants to replace the reference data for some reason (e.g. the data are too detailed or the user wants to merge his thematic data with other thematic data that does not use the same reference), the thematic data have to be processed in order to maintain data

K. Jaara (✉) · C. Duchêne
IGN France—COGIT Laboratory, Saint-Mandé cedex, France
e-mail: Kusay.jaara@ign.fr

A. Ruas
IFSTTAR, Champs sur Marne, France
e-mail: Anne.ruas@ifsttar.fr

consistency. We call this processing thematic data migration (Jaara et al. 2012). If the initial relative position is not taken into account, the process is called thematic data integration. Relative position is described by relations. Relations have been used previously to adjust objects and to improve quality (Wallgrün 2012).

Thematic data migration can be used to propagate thematic data between different levels of detail (LOD) of multi representation database (MRDB). Independently of MRDB, migration can be used when a lower LOD dataset has to be derived by generalisation from an initial dataset that is a combination of topographic and thematic data, a possible process is to generalise the topographic data then to migrate the thematic data. In either case when the initial and final LOD are very different. Thematic data migration may not be sufficient, and thematic data generalisation might be needed in order to get a consistent result. Thematic data generalisation is not studied in this paper.

Jaara et al. (2012) proposed a first workflow for thematic data migration (Fig. 1); initial spatial relations between thematic and topographic data are
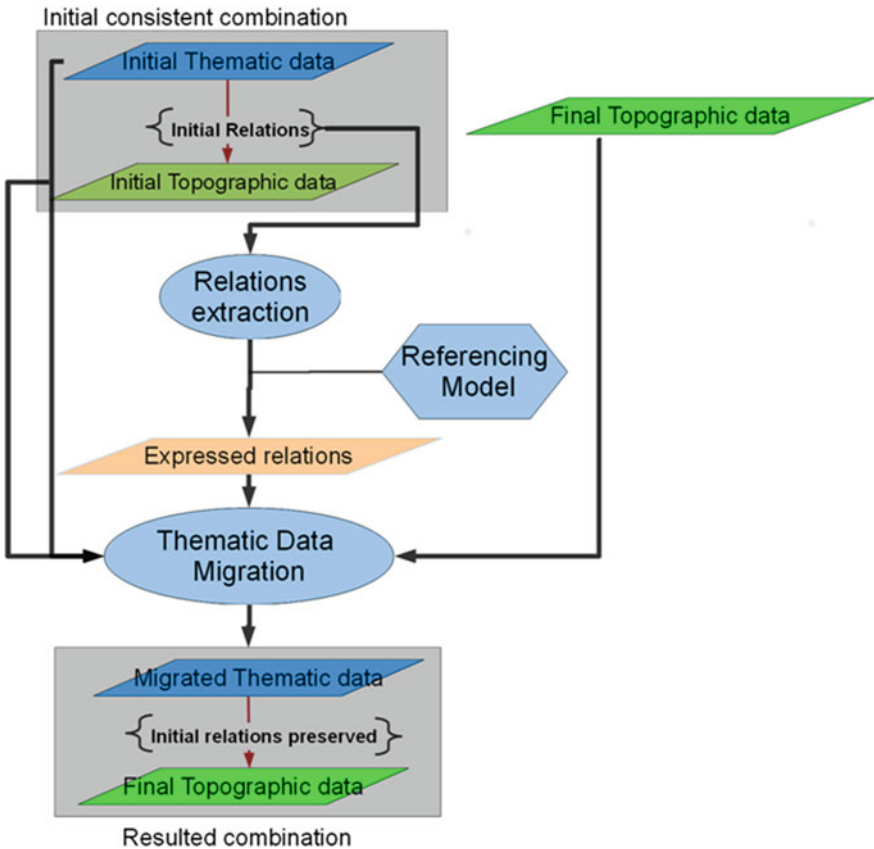
Fig. 1 General workflow of the thematic data migration process proposed in (Jaara et al. 2012)

extracted and stored in a dedicated "referencing model" also introduced in the paper. Then a migration process takes place, which seeks for processing the initial spatial relations.

This paper proposes an updated version of thematic data migration process, which aims at considering the case where the relations between thematic and topographic data have to be modified between the initial and the final state.

Moreover, this paper presents more in detail our model to describe the relations, taking into account the case where they are modified, as well as a detailed description of the method used to relocalise the thematic data on the topographic data guided by the described relations.

The paper first introduce the idea of relation modification, proposes a formulation of the studied problem, and a new version of the general thematic data migration workflow is proposed in Sect. 2. Section 3 describes the modelling of relations and their modifications. In Sect. 4, the thematic data relocalisation based on relations is detailed.

The approach is illustrated by a use case on road accidents, for which obtained results from the implementation of our workflow are shown in Sect. 5, and results obtained with the implementation of our method are shown.

Introduction of the running example used throughout the chapter:

The two following running examples will be used in the paper. Both consist in road accidents (thematic data) located on roads (topographic data), that should be migrated on a lower LOD road dataset.

In the first example, two accidents exist, one is close to the roundabout and the other is further (Fig. 2a *top*). The roundabout is represented by a crossroads in the final reference data (Fig. 3c, *top*). If the accidents are relocated to the nearest road point in the final topographic database, the relation to the end of the road will be lost (Fig. 2c, *top*), although it may be an important relation for accident analysis.

In the second example (Fig. 2a, *bottom*), we have an accident in front of a building and next to a river–road intersection; the building is not in the same relative position in the final reference database (Fig. 2b, *bottom*), if the accident is relocated to the nearest point of the road on the final topographic base. The accident will no longer be in front of the building, although it could be an important relation for the user.

2 Modification of Spatial Relations Between Thematic and Topographic Data

2.1 Why do some Relationships have to be Modified?

The representation of the real world is not the same, from one database to another, the difference could concern the size, location and shape. Objects could be aggregated with others or deleted; the dimension of their geometry may change,
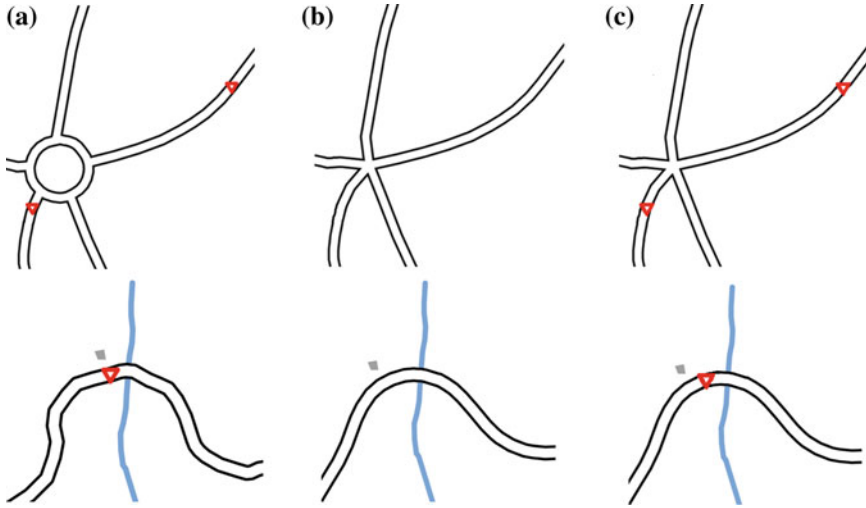
Fig. 2 Thematic data migration, use case of accidents and roads. **a** Initial combination of accidents with initial topographic database, **b** final topographic database, **c** thematic data migration result using nearest position

e.g. a river can be represented by a surface or a line. These changes modify the spatial relations between the initial thematic data and the initial topographic data. For instance, the initial proximity relation with the roundabout in Fig. 2 will necessary change, because the roundabout is represented as a crossroads in the final reference data. Relations may need to be modified even if the reference object is not modified, for example, a disjoint relation between a thematic region and a reference region might be modified into touch if the final reference data has smaller scale and the distance is small enough.

2.2 Formal Problem Statement

Topographic objects are the reference for thematic objects. One thematic object can have several relations with topographic objects. Conversely, one topographic object can have relations with more than one thematic object. We consider that each relation $r \in \mathcal{R}$ involves only two objects, one object is thematic a and the other is topographic b.

The objective is to extract thematic data relations characterized by needed attributes in order to use these relations to localize the thematic data on the final reference data. In other words if a, a' are the initial and final thematic objects respectively, and b, b' are the initial and final topographic objects respectively and $\mathcal{R}, \mathcal{R}'$ are the initial and final relations then:

$$a\mathcal{R}b \wedge b \rightarrow b' \Rightarrow ?\,\mathcal{R}'\,? \ominus a'/a'\mathcal{R}'b'.$$

This means, when topographic data b is changed into b′, the first task is to derive the expected relations of the final state $\mathcal{R}'$ changed from $\mathcal{R}$, then to obtain thematic data a′ that satisfies the $\mathcal{R}'$ relations.

3 General Workflow of Thematic Data Migration When Relations can be Modified

In Fig. 3 we propose a developed version of the data migration workflow of Jaara et al. (2012), where we integrate the phase of relation modification.

According to the workflow, thematic data migration consists of:

1. *Relation extraction*. Significant types of relations are identified depending on the considered application case, then relevant instances of these relations are extracted from initial data (e.g. accident a1 is on road r1 and accident a2 is close to the roundabout ra1).
2. *Matching*. Initial and final topographic data are matched to detect the corresponding objects and changes (e.g. the roundabout with the crossroads). Any data matching process can be used if it works on the considered topographic data. For our running example, we need a road network matching process like Mustière and Devogele (2008).
3. *Relations modification*. Expected relations in final data are inferred, i.e. relations corresponding to initial relations are identified using final topographic objects, and how they should be modified is also identified when needed (e.g. proximity relation with roundabout becomes a "hosted by" inclusion relation with the crossroad).
4. *Thematic data relocalisation*. The expected relations in final data are used to control the relocalisation process.
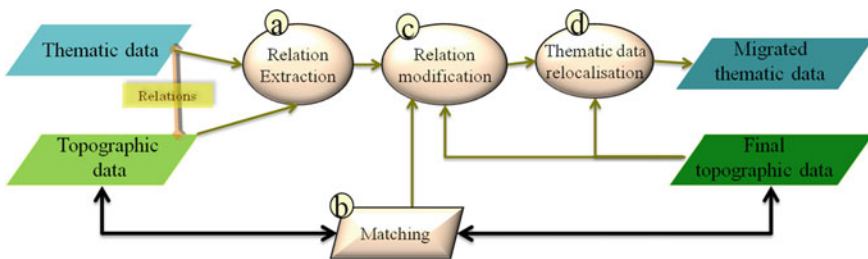
Fig. 3 Proposed workflow for thematic data migration

4 Modelling of Spatial Relations and Their Expected Modifications

4.1 What are the Relations to be Extracted?

This section gives an overview of the considered types of relations and existing works on their modelling, and then focuses on the specific nature of thematic-reference relations.

Many classifications of spatial relations exist. In Ruas (1999) they are classified as topologic, angularity and neighbourhood relations. For Jones (1997), relations are classified as topologic or directional or proximity relations. In the model by Jaara et al. (2012) they are classified into three categories: topologic, metric (e.g. proximity, orientation) and other (e.g. local shape of road beside an accident).

Regarding topologic relations, the most famous model is the 9 intersections model by Egenhofer (1991) for regions, extended for line–line and line–region by Egenhofer and Herring (1990). More recently, Wallgrün (2012) further studied their mathematical formulation. Examples of models for directional relations are those by Papadias and Theodoridis (1997) for regions, Kurata and Shi (2009) for line–region. Proximity is usually represented by means of a distance. Among "other" relations, relations defined by Mathet (2000) like go round can be found, e.g. a river goes round a city.

If one of the two objects is thematic, we notice three features:

- The first is that the importance of relations is linked to the nature of thematic objects, which means that some relations are semantically more important for the given thematic data. For example, if there are two relations (boat accident—reference road) and (boat accident—reference river), the boat accident is more related to the river than to the road, so the relation to a road is less important.
- The second feature is the special relation of "hosted by", which we define as a particular topologic relation of inclusion between the thematic object and a topographic object, which has a semantic relation with it. For instance road accidents are hosted by roads, so they only exist on roads.
- The third feature is the use of characteristic objects, which are identified as additional objects that help to get a better description of the thematic data position. Characteristic objects are extracted in both initial and final reference data, and then they are matched. For example a roundabout is a characteristic object that could be extracted from the initial reference database and then matched with the crossroads characteristic object of the final reference database.

4.2 Relation Extraction and Modelling

The relations to be extracted are chosen according to the use case and user requirements. These are modelled and stored by using a referencing model. In

Touya et al. (2012), a model of ontology for relations and relational constraints has been proposed in order to unify the definitions of relations and constraints for automatic generalisation. From the proposed model (shown in Fig. 4), we use the following elements:

- Spatial relations are either quantitative (e.g. distance or orientation) or binary (e.g. topological relations).
- Binary relations are extracted by examining certain conditions, and quantitative relations have conditions of relevance.
- A relation has properties. For example, the proximity relation between a point and a building could be represented by the distance property between the point and the building border.
- A relation has members which are the related objects.
- Relations are detected by dedicated methods that also detect their properties.
- Binary relations may have a number of predicates, every predicate is a separated relation, and the whole is a family of relations so that two objects cannot have two relations from the family. For example, in the 4 intersections model by Egenhofer et al. (1991), we consider a family of 8 relations: included, includes, covered by, covers, overlaps, equals, meets, disjoint.
- In some cases it is unclear if a given binary relation is achieved or not (e.g. a building is almost parallel to a road), or it is unclear which of two neighbor relations of a family is achieved (e.g. very small distances between regions, so there can be doubt, is it "disjoint" or "touch", because of perception thresholds). The term "fuzzy relation" is used for that (fuzzy parallelism; fuzzy disjoint touch).
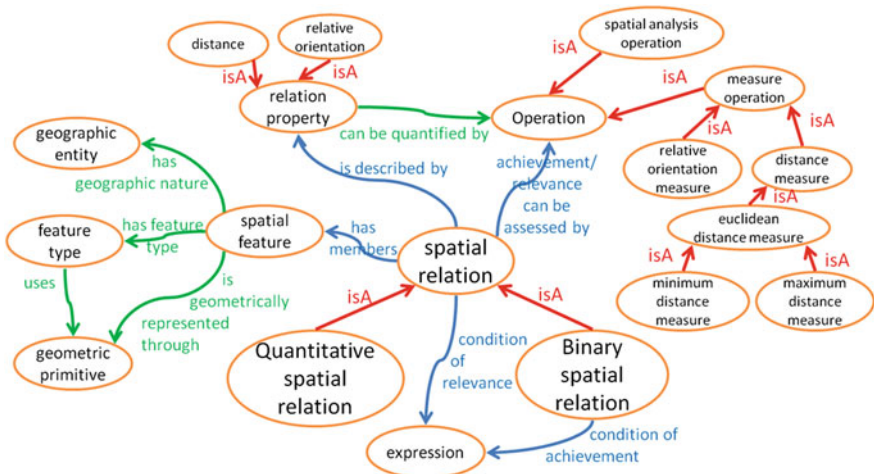
Fig. 4 Model for spatial relations proposed in Touya et al. (2012)
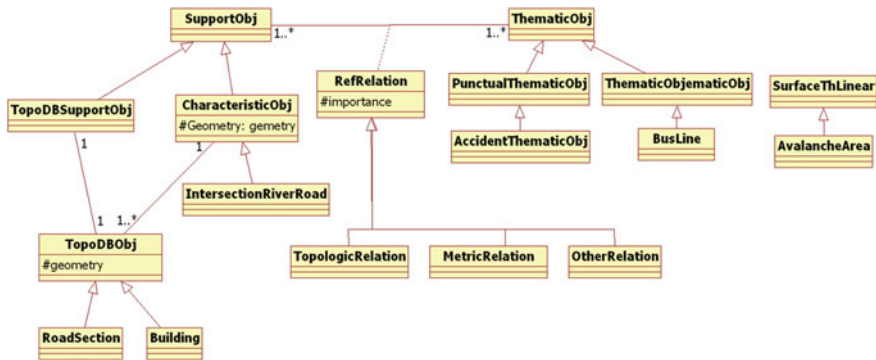
Fig. 5 Class diagram of the referencing model of Jaara et al. (2012)

In the model presented in Jaara et al. (2012), relations are not only between thematic and reference objects. As shown in Fig. 5, relations could be with topographic objects or with what we introduced as being characteristic objects (see also *4.1*). Every object among the topographic objects and the characteristic objects that are involved in a relation with a thematic object is called a support object.

The proposed ontology by Touya et al. (2012) shown in Fig. 4 has been extended in order to include the thematic data referencing model. Spatial feature can be either characteristic or topographic or thematic. Every relation has two members, member1 which is thematic and member2 which is either characteristic or topographic object. Figure 6 shows the modifications and how the ontology has been used to represent a case of road accidents.
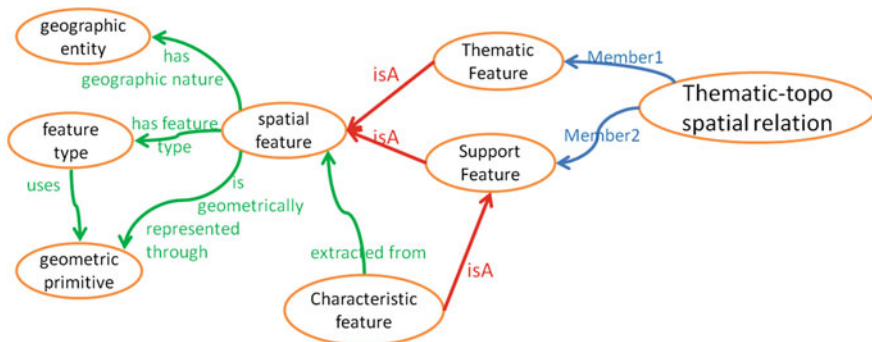
Fig. 6 Extension of relations ontology (Touya et al. 2012)

4.3 Modification of Relations According to the Final Data

The needed modifications of relations have to be identified depending on the matching between initial and final topographic data. The matching result could be the disappearance of topographic object, change of size, change of shape (e.g. Building aggregated with others), difference of representation (e.g. from surface to a line), change of geometric position (Fig. 2 second example) between the initial and the final topographic database. According to the matching results, some relations have to be adjusted by changing their attributes (e.g. distance from the building if the building is bigger in the final topographic database), others are transformed to other types of relations (e.g. initial relation with a river surface if it is represented by a curve in the final topographic database). Relations could be completely ignored, e.g. the relation with a building that is removed in the final topographic database.

The modification of relations is guided by rules. Rules have to be defined for every type of relation according to the matching result. There is no need to know the final topographic database to lay down the rules. Rules may also be related to the message of the map. Modification rules for topologic relations can be defined with the help of conceptual neighborhood graphs (Egenhofer and Mark 1995). For example, one of the rules that could be defined according to the conceptual neighborhood graphs (Fig. 7).

If the initial relation is (r = r3) and the distance between the region and nearest point of road in the final topographic database is less than 0.5 cm, the r relation will be changed into (r' = r5).

4.4 Need for an Evaluation of Relation Satisfaction

After identifying the expected modification of relations, the objective is to find the best position for the thematic object in the final database. For this reason, we need a way to evaluate a given position of thematic data in the final reference data. The

Fig. 7 Part of the conceptual neighbourhoods of line–region relations (Egenhofer and Mark 1995)
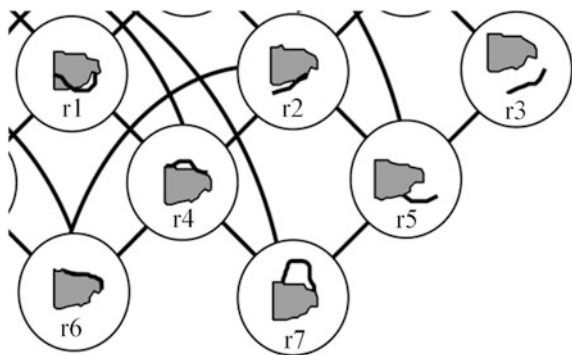
evaluation shows the degree of respect of the expected final relations. We introduce the relation satisfaction measure, which is a method of quantifying how much a relation is fulfilled by a given pair of objects. The method depends on the relation type and attributes. For instance, for the proximity relation between an accident and a building, we can define a satisfaction measure for this relation in the following way:

$$
\begin{array}{ll}
\textit{If } |d' - d| >= 100 \text{ m} & S = 0 \\
\textit{If } |d' - d| <= 30 \text{ m} & S = 1 \\
\textit{Otherwise} & S = 1 - ((|d1 - d2| - 30)/70)
\end{array}
$$

where d is the expected distance and d' is the distance of a given location to evaluate.

In binary relations, the neighbourhood graph can help detect the level of satisfaction. If we take the case of line–region topologic relation (Fig. 4),

$$
\begin{array}{ll}
\textit{If the same relation exists for the given position} & S = 1 \\
\textit{If the distance is one transition (e.g. from R6 to R4)} & S = 0.5 \\
\textit{Otherwise} & S = 0.
\end{array}
$$

5 Thematic Data Relocalisation Based on Relations

After the expected relations between thematic and topographic data in the final database have been computed, we search the location that best satisfies these relations. This relocalisation process is an optimisation issue. Our solution is a local search around the initial position, based on the discretisation of possible space. It consists of two steps:

1. Detect possible positions for each thematic object in the final reference database: if a thematic object is held by a reference object, and if the reference object is a surface or a line, it is discretized to get a number of points. The resulting point set is intersected with a buffer around the initial thematic position to refuse far positions. If there is no holding reference object the buffer around the initial location of the thematic object will be discretized to obtain the possible positions.
2. Choose the best location based on satisfaction measures using a multi-criteria decision system: the level of satisfaction of every relation is computed for every possible location resulting from the discretisation operated at stage 1. We consider the search for optimal location based on expected relations as a multi-criteria decision problem, where satisfying of relations are the criteria. Indeed it is not always possible to satisfy all relations. We notice that it is better to ignore some relations and satisfy others at best, rather than trying to find a best compromise among all of them. After a study of existing multi-criteria decision

methods, we decided to use PROMETHEE II (Brans and Mareschal 2005) as a decision system because it enables to favour solutions according to criteria.

PROMETHEE II (Preference Ranking Organization Method for Enrichment Evaluations) is a multi-criteria decision method classified as partial aggregation approach, which means that it works by comparing each pair of possible solutions. The result of using the decision method is obtained by ranking the possible positions, ordered according to how much they satisfy the expected relations.

3. Refine the neighbouring area around the best location found in stage 2: in this stage, the neighbouring area is discretized with a smaller equidistance, to test if there is a better solution with a higher ranking that was skipped by the first large equidistance.

6 Application of the Workflow on the Use Case of Accidents and Roads

Our method has been implemented in Cartagen platform (Renard et al. 2010). This section illustrates the method on the running example of Fig. 2 (reproduced hereafter in Fig. 8), and shows the obtained result.

The thematic data migration is applied as follows:

1. *Extraction of characteristic objects*. As shown in Fig. 9, the roundabouts are recognised as characteristic objects. They are extracted from the road network by looking for round faces in the graph which are derived from the road network using the miller index proposed by Sheeren (2005). In the second example, the characteristic object is the river–road intersection, which is extracted by using simple line intersection.

2. *Relation extraction*. First, we need to find the road that is hosting the accident, then to explore the neighbourhood of every accident to detect every possible support object. In example 1, the roundabout is found as a characteristic object because it is not far and is connected to the same host road, in example 2 the intersection of river–road and the building are found as possible support objects. The next step is to extract all the possible relations between the accidents and the possible support objects. All conditions of relations relevance are tested. The relations are as follows:

   - point–line topological relation of "hosted by" between accidents and roads, in order to extract the host road;
   - point–point topological relation and orientation relations between accidents and buildings;
   - point–point proximity relations between accidents and punctual characteristic points;
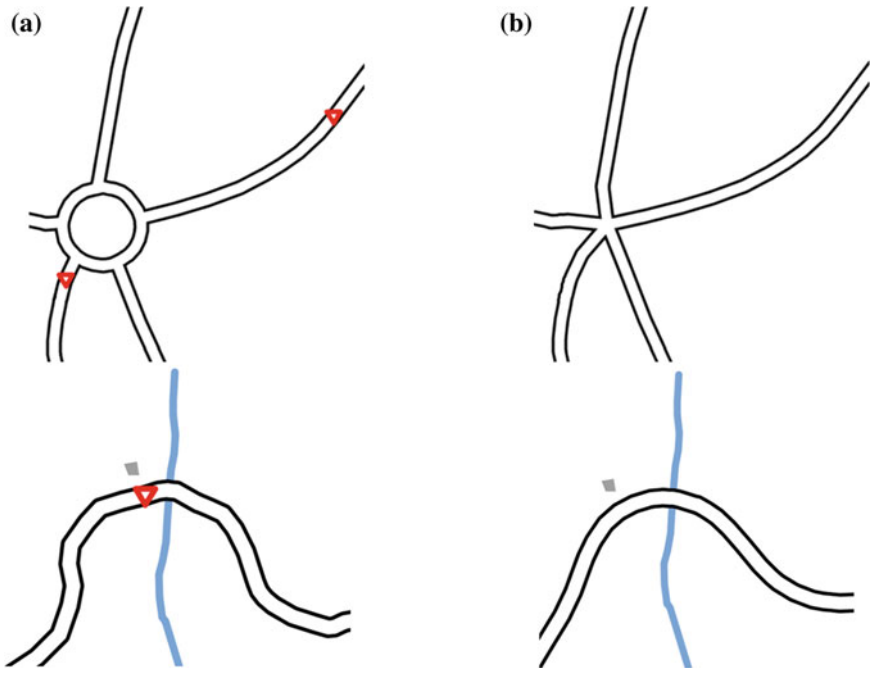
Fig. 8 Initial and final reference data and the initial location of accidents. **a** Initial combination.
b Final topographic database
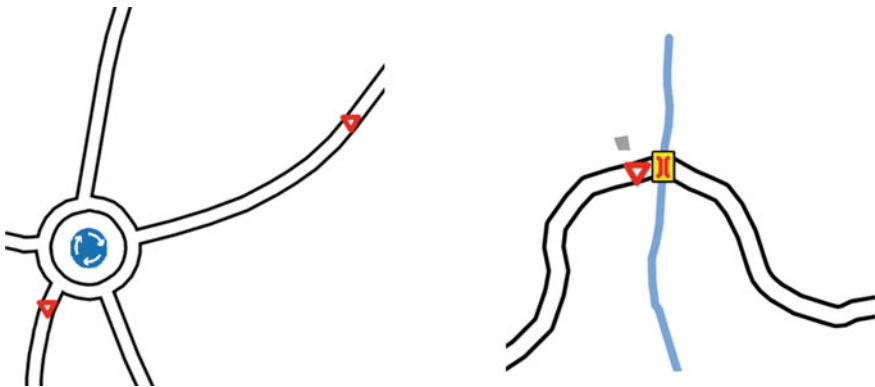
Fig. 9 Extraction of the initial characteristic objects

- point–roundabout proximity relations between accidents and roundabouts that are connected to their host road.

  In example 1, the following relations were extracted

- two hosted by relations were extracted, one for every accident;
- two proximity relations of point–roundabout were extracted, one for every accident.

  In example 2, the following relations where extracted

- two hosted by relations were extracted, one for every accident;
- one point–point proximity relation with the river–road intersection;
- one orientation relation between the accident and the building;
- one proximity relation with the same building.

3. *Extraction of final characteristic objects*. This means the crossroads in the first example and the river–road intersection in the second example (Fig. 10).
4. *Data matching*. First, the change of reference topographic data was detected data matching. In our implementation, we use the automatic method of by Mustière and Devogele (2008) to match the road network and find the corresponding roads in both examples. Another matching process is also done between buildings. Secondly, the change of characteristic objects between the initial and the final database were detected. This corresponds to the link between the roundabout and the crossroads in the first example and the link between the two river–road intersections in the second example.
5. *Relation modification*. In the first example, the roundabout becomes a crossroads. In the configuration of point–roundabout proximity relation, if the roundabout becomes a crossroads, the relation becomes point–point. Moreover, in this case, the resulting map is dedicated to the study of road accidents so that if an accident is too close from a junction (50 m here), it should be considered hosted by the junction. This is the case of accident a1 here, which is initially closer than 50 m to the roundabout. Therefore, its final expected relation with
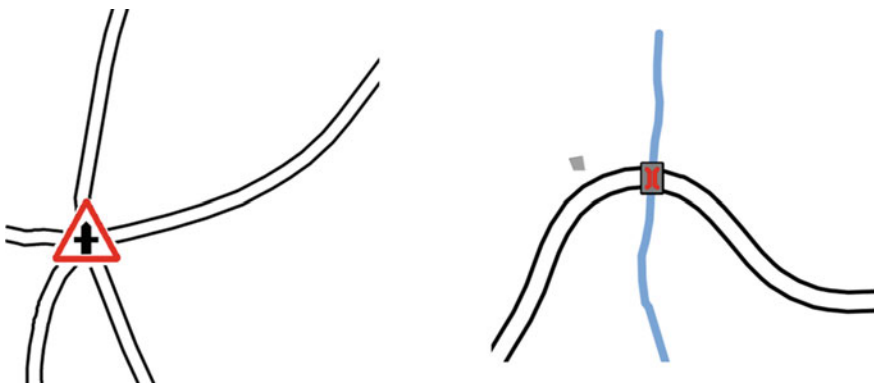
Fig. 10 Extraction of final characteristic objects
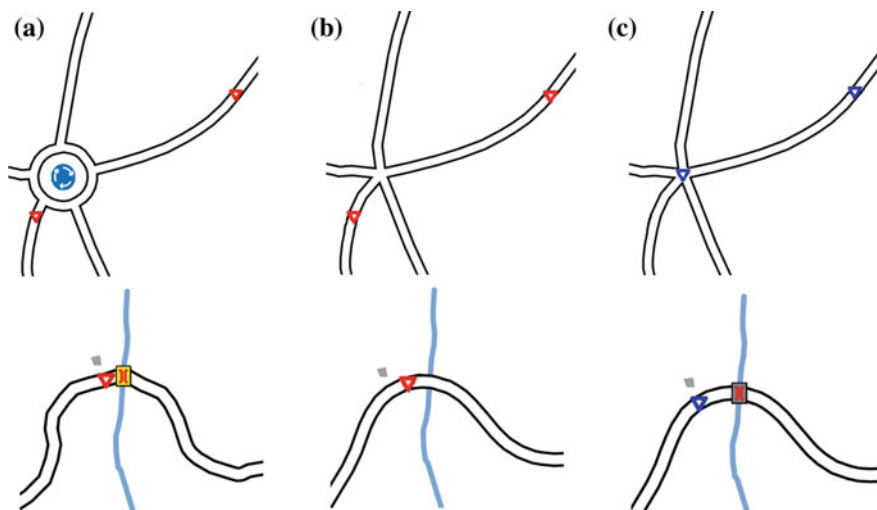
Fig. 11 Results of thematic data migration. **a** Initial combination of accidents with initial topographic database. **b** Final combination with thematic data migration to nearest road point. **c** Final combination using our method of thematic data migration

the crossroads is "hosted by". Accident a2, further than 50 m from the roundabout, will keep a proximity relation with the crossroads, except that the distance will be equal to the initial distance to the centre of roundabout. In the second example, relations will stay the same. Note that for another use, like a real-time navigation map, it might have been needed to increase the distance from a1 to the crossroads instead of putting it on the crossroads, to make it clear that not the whole roundabout is blocked by the accident.

6. *Thematic data localisation*. The final holding road is discretized to a point every 5 m not further than 300 m from the initial accident location. The preferred relations are evaluated for every possible location and evaluation values are injected into the mutli-criteria system. Then the system gives us the best location. Another discretisation of 1 m is done to get the precise location.

Figure 11 presents the obtained result, which corresponds to what was expected.

7 Conclusion

The paper is follow up of the research on thematic data migration, first presented by Jaara et al. (2012), where model of thematic data migration was proposed.

The idea of relation modification was introduced to take into account the modification of reference data and the new level of detail. A modified workflow

for thematic data migration is presented taking into account the modification of relations. The proposed relocalisation process is based on the multi-criteria decision method where the criteria are the satisfaction level of relations are the criteria. The whole workflow is illustrated by a use case of road accidents.

The proposed workflow will have to be tested on real data, and it is important to study the effect of different parameters. In addition, the workflow needs to be adapted for the case of linear thematic data. Formulated relations can be used to guide the process of topographic data generalisation. Relations can also be used to constraint thematic data generalisation like the aggregation of a group of accidents around a roundabout into a unique punctual thematic object.

References

Brans JP, Mareschal B (2005) PROMETHEE methods. In: Figueira J, Greco S, Ehrgott M (eds) Multiple criteria decision analysis: state of the art surveys. Springer, London, pp 163–196

Egenhofer M (1991) Reasoning about Binary Topological Relations. Second Symposium on Large Spatial Databases, Zurich, Switzerland, Gunther O, Schek H-J (eds.), Lecture Notes in Computer Science, Aug 1991, 525, Springer-Verlag, pp. 143–160

Egenhofer M, Herring J (1990) Categorizing binary topological relationships between regions, lines, and points in geographic databases. Technical report 90-12, National Center for Geographic Information and Analysis, University of California, Santa Barbara

Egenhofer M, Mark D (1995) Modeling conceptual neighborhoods of topological line–region relations. Int J Geogr Inf Syst 9(5):555–565

Jaara K, Duchêne C, Ruas A (2012) A model for preserving the consistency between topographic and thematic layers throughout data migration. In: 15th international symposium on spatial data handling, Bonn

Jones C (1997) Geographic information systems and computer cartography. Longman, Harlow

Kurata Y, Shi H (2009) Toward heterogeneous cardinal direction calculus. In: Mertsching B, Hund M, Aziz Z (eds) KI, Lecture notes in artificial intelligence 5803, Paderborn, Sept 2009, pp 450–457

Mathet Y (2000) Étude de l'expression en langue de l'Espace et du Déplacement: analyse linguistique, modélisation cognitive, et leur expérimentation informatique. Thèse de Doctorat, Université de Caen, France

Mustière S, Devogele T (2008) Matching networks with different levels of detail. GeoInformatica 12(4):435–453

Papadias D, Theodoridis Y (1997) Spatial relations, minimum bounding rectangles, and spatial data structures. Int J Geogr Inf Sci 11:111–138

Renard J, Gaffuri J, Duchêne C (2010) Capitalisation problem in research—example of a new platform for generalisation: CartAGen. In: 12th ICA workshop on generalisation and multiple representation, Zürich

Ruas A (1999) Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. Thèse de doctorat, Université de Marne-la-Vallée, France

Sheeren D (2005) Méthodologie d'évaluation de la coherence interrepresentation pour l'integration de bases de données spatiales. Phd thesis, Université Paris 6, France

Touya G, Balley S, Duchêne C, Jaara K, Regnauld N, Gould N (2012) Towards an ontology of generalisation constraints and spatial relations. In: 15th ICA workshop on generalisation and multiple representation, Istanbul

Wallgrûn J (2012) Topological adjustment of polygonal data. In: Proceeding of 15th international symposium on spatial data handling, Bonn, pp 187–201

A Novel Approach of Selecting Arterial Road Network for Route Planning Purpose

Hongchao Fan, Hongbo Gong and Qing Fu

Abstract The most of existing algorithms for road network selection are proposed for the visualization purpose. Hence, the connectivity of road network for route planning has rarely been considered in the previous works. In this chapter, we propose a novel method of road selection, whereby decisive paths that distinguish the suboptimal route from the optimal one can be identified and added to the high-layer network which is formed mainly by the connectivity of the crucial cities. This benefits the improvement of vertical partitioning and finally the construction of a high-layer road network that allows the optimal route planning. A case study in Bavaria State, Germany, reveals the feasibility of the proposed approach.

Keywords Generalization · Road network · Selection · Route-planning

1 Introduction

In cartography, selection is considered as an important operation for the model generalization (Mackaness 2007). It consists in choosing the relevant information in relation to the target database specification. In traditional cartography, the selection of roads is merely dependent on their relative importance in terms of functional classes, administrative levels, widths etc. However, the results of this kind of approach may contain some unacceptable artifacts. For instance, the road

H. Fan (✉)
Department of GIScience, University of Heidelberg, Heidelberg, Germany
e-mail: hongchao.fan@geog.uni-heidelberg.de

H. Gong
Kotei Navigation Co. Ldt, Wuhan, China

Q. Fu
College of Survying and GeoInformatics, Tongji University, Shanghai, China
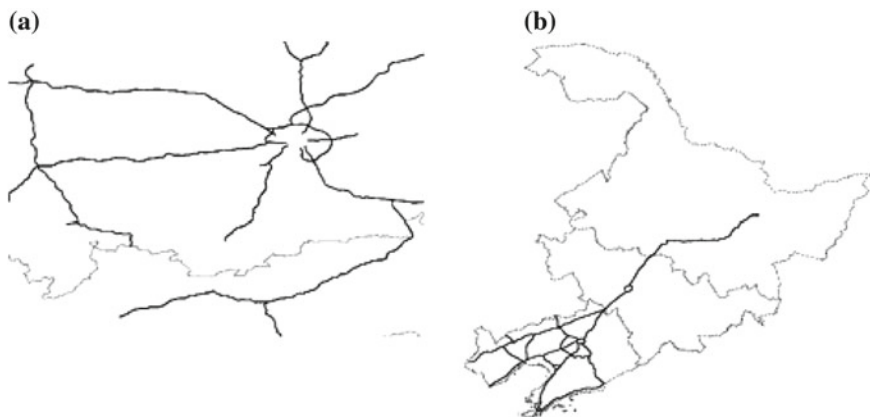
Fig. 1 Unacceptable results of selection based on insufficient governing information. **a** Disconnected road network in southern Germany and western Austria. **b** Shrewd density distribution of selected roads in Northeast China

network in southern Germany and western Austria becomes disconnected in many places including the border region after selection operation (Fig. 1a), while the road network in Northeast China reveals an unbalanced distribution after the selection process (Fig. 1b) and the reader might get the impression as if the north part of the region were uncultivated, although in the reality there is a well-developed traffic infrastructure.

This chapter presents a road network selection process with the attempt to close the aforementioned research gap. The work is focused on selection of road network in the high-layer of navigation data, in which the road network is selected for (very) long distance route planning and displayed at small scale of inter-city level.

The rest of the chapter is structured as follows: Sect. 2 gives a review of the existing researches about road network selection; Sect. 3 presents the proposed approach with description of concept and work procedure; then a case study and its evaluation is described in Sect. 4; finally the work is concluded in Sect. 5.

2 Related Works

The process of selection in cartography consists in choosing the relevant information in relation to the target database specification. Therefore, it is often used for deriving abstractive road network from more detailed data, which are used for route planning at different levels. In the past decades, a number of algorithms have made available for road network selection. They are approached using various theories and measures, which can be classified as follows:

Semantics: Li and Choi (2002) discussed the association of road elimination with six types of thematic attributes: type, length, width, number of lanes, number

of traffic ways and connectivity. Sinha and Flewelling (2002) explained a method of multi-criteria line generalization to incorporate road category and road class into the generalization process. Kulik et al. (2005) realized an ontology-driven map generalization algorithm, called DMin, to simplify a road network in which the road classes and tasks are parameterized as semantic weights.

Graph theory: Road selection can be based on graph theory, such as directed network (weighted graph), minimum spanning tree, dynamic decision tree, shortest path algorithm, and so on. Mackaness and Beard (1993) and Mackaness (1995) utilized directed network (weighted graph) to support road-network selection, and derived several preliminary rules for the generalization process from graph theory. In particular, Thomson and Richardson (1995) as well as Thomson and Brooks (2000) used the concept of minimum spanning trees for road-network selection. Peng and Muller (1996) introduced a dynamic decision tree structure in an attempt to partly circumvent the problem of urban road-network generalization through the use of object classification and aggregation hierarchies, topological data structure, decision rules, and artificial intelligence technology. Touya (2007) used shortest paths algorithm (Dijkstra 1959) in rural road network between important points to guide the road selection and check the continuity as in urban/rural interface zones. The shortest path algorithm in this context indicates the relative importance of road links and nodes.

Stroke: Thomson and Richardson (1999) defined a structural element "stroke" which can be derived following the "good continuation" principle in Gestalt psychology (Wertheimer 1938; Thom 2006) approached the functional-graphical nature of strokes via a broader consideration of perceptual grouping. A stroke is a group of roads gathered by not only continuous curvature but also homogeneous semantic attributes and is stopped at hanging node, roundabout, branching crossroad or turning. Characteristics of a road network can be preserved if salient strokes are kept during data reduction or selection process. Thom (2005) used it to detect dual carriageways; Heinzle et al. (2005) treated stroke as a pattern of road network. Edwardes and Mackaness (2000), Touya (2007), Chaudhry and Mackaness (2005), and Liu et al. (2009) applied the methods of stroke to reach the good results road-network generalization.

Road density: Road density is another useful constraint or measure for the selection as it provides metric and statistical information about overall road distribution at the macro and micro levels. Zhang (2004) presented a method to select salient roads based on connection analysis which is a favored criterion for maintaining the density variations in road network. Zhang et al. (2008) developed an object-oriented measurement to measure the distribution of density across a map in generalization. Liu et al. (2009) proposed an algorithm for road density (grid method) analysis based on skeleton partitioning for road generalization. Mesh density is another measure indicating the local distribution density of road network. Mesh was defined as the smallest sub-region which could not be divided by roads. Hu et al. (2007) put forward this idea and presented a method for selective omission of street network for digital map generalization. Then Chen et al. (2009) presented an approach to automatic map generalization by using mesh density. Tian (2008) presented a generalization algorithm based on the ordered

generalization tree structure to support the progressive representation models. In these research works, the road-network generalization is constrained by the topological and distributional information such as minimum separation rule, connection analysis, progressive representation, skeleton partition and mesh integration.

Pattern: Patterns were defined as property within objects, or between objects that is repeated with sufficient regularity by Mackaness and Edwards (2002) who argued that it is useful to consider any given map as a view of the subset of all possible patterns inherent among objects in the database, and viewed the process of generalization as being about manipulation and portrayal of pattern for any given scale and theme such as road network. Early in 1995, Sester identified in her dissertation nine node types in the low level of details and several road types according to their node degrees and the arrangement of the intersecting lines. Urban pattern was studied by many researches. Heinzle et al. (2005) conducted the automatic localization of a city centre on the basis of pattern analysis and detection in road networks. They defined four kinds of patterns: strokes-pattern, star-like pattern, ring-like pattern and grid-like pattern. In their follow-up works, methods were developed to automatically recognize these patterns (Heinzle et al. 2006; Anders 2006; Heinzle and Anders 2007; Heinzle et al. 2007). Touya (2007) presented a generic process for road-network selection by enriching the data with structures and patterns recognition. Zhang (2004) described the components and properties of grid-like patterns, star-like patterns for road-network modeling and generalization using geometrical parameters and objects.

Area ratio of road symbols: Gulgen and Gokgoz (2008) developed a road-selection method which can preserve the area ratio of road symbols in a map. The areas in the map space covered by road symbols were used as a measure to decide the maximum reduction limit. Exaggerated road symbols in target scales claim more area, thus requires the elimination of some unimportant roads. The method aims to show cartographic relationship between retained roads and scale change during the generalization.

In general, almost all the above mentioned existing works have been dedicated for the visualization purpose. They have been developed with the full consideration of the map legibility at different scales. There are rare approaches with the aim of route planning in car navigation system.

3 The Algorithm Based on Optimal Path Comparison and Validation

The proposed algorithm is named Optimal Path Comparison and Validation (OPCV) algorithm in this work. In this section the algorithm is described in detail, whereby it is termed using its abbreviation in the following sections.

3.1 The Concept of OPCV Algorithm

Figure 2 shows the typical shortest way-finding problem when the optimal routing algorithm such as the accelerating A* algorithm is performed on a network partitioning based on an unreasonable hierarchical road network. Since a middle-layer road marked in grey in Fig. 2a is not included in high layer, the accelerating A* algorithm will therefore follow the high-layer roads in red, thus provide a longer route in the end. In order to support accelerating A* algorithm to obtain the correct optimal route in green in Fig. 2b, the inclusion of the necessary middle-layer road in the high-layer road network will lead to a more reasonable hierarchical road network. This is the task of an Optimal Path Comparison and Validation (OPCV) algorithm.

The idea of the OPCV algorithm is inspired by a system test method for the examination of the route-planning function of a navigation unit. A routing task typically involves a given start point, destination and a route-planning strategy. It can be tested in two environments with different hardware, datasets and algorithms as shown in Fig. 3. The correct results can be definitely calculated with A* algorithm working on a single layer of road network as proved by Hart et al. (1968). But the accelerating A* algorithm that aims at speeding up the searching process, thus works with network partitioning cannot ensure the correctness of results, depending on how the hierarchy of the road network is established. If the results of two algorithms are different, there must be something wrong with the network partitioning for the accelerating A* algorithm.

Theoretically, this test method can be used to uncover problems for any arbitrarily given routing task, but obviously it is inefficient to calculate the paths with so huge road links from the low layer. In order to increase the efficiency, a simulation environment on test server for a navigation unit with simplified databases can be created. Since only some necessary roads from the middle layer will be included to the high layer and it seldom happens that a road from the highest layer is connected with a road from the lowest layer, the roads from the low layer can be eliminated from the single-layer road network for A* Algorithm, while the
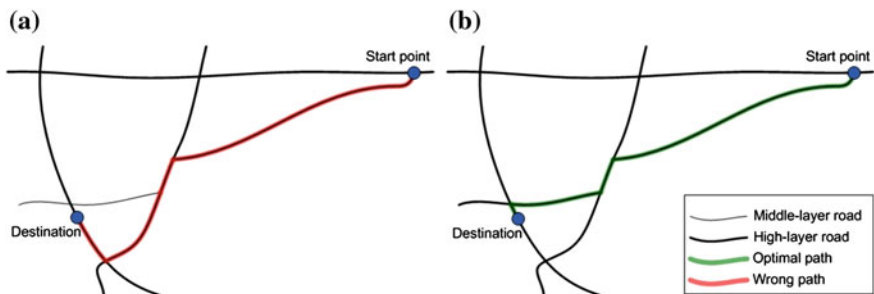
Fig. 2 Typical problem caused by unreasonable hierarchical road network. **a** Suboptimal path when a necessary road is not included in high layer; **b** optimal path when the road is picked up into high layer
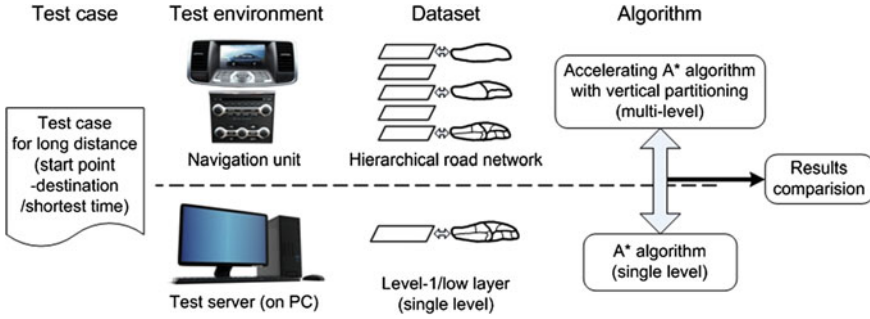
Fig. 3 System test of a navigation unit

hierarchical road network for the accelerating A* Algorithm will just consist of middle layer and high layer. In the test, all nodes in the test region and the nodes of the links outside the test region that are directly connected with can be selected as valid terminating points for the route calculation.

Figure 4 illustrates the automated test process of the OPCV algorithm for long-distance route planning and the shortest time as the route-planning strategy. The process is composed of the following steps:

Step 1: Calculate all optimal routes with A* algorithm based on single-layer road network (middle-layer road network) which can be built up using the selection methods by individual attributes, road priority and road connectivity.

Step 2: Calculate routes with accelerating A* algorithm (network partitioning) based on hierarchical road network with two layers (middle layer and high layer). The high-layer road network is built up using the selection method by important cities. Compared with A* algorithm, this method speeds up the process, but creates a small proportion of wrong or suboptimal routes because of the suboptimal hierarchical road network.

Step 3: Compare the results from two algorithms, and recognize the differences where wrong or suboptimal routes exist.

Step 4: Analyze the problems and adjust the hierarchical road network by including the necessary middle-layer roads to high layer. Road links in the road network may interfere with each other. After a road link from the middle layer is upgraded to the high layer, some other middle-layer road links may be modified accordingly.

Step 5: Calculate the paths again with accelerating A* algorithm based on the adjusted hierarchical road network to verify and validate if all new routes are identical to the results from A* algorithm. The experimental results reveal that most optimal routes can be successfully verified, a few needs one or more iterations.
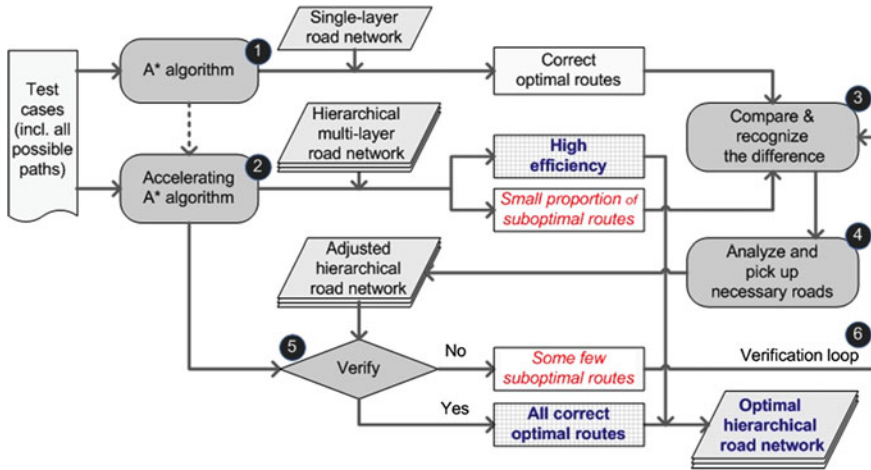
Fig. 4 Process of the OPCV algorithm. (Undesired results in *red* and expected results in *blue*)

3.2 Initialization of OPCV Algorithm for Testing

The initialization of OPCV algorithm for testing aims to get the system prepared for the traverse of all possible paths in a test region (usually a province or a state) going through given start points and destinations. Two point sets are necessary for the configuration: one containing all nodes within the test region including the nodes on its border; the other containing some specified nodes outside the test region.

Route planning for long-distance travel is a research focus of the OPCV algorithm. Generally all paths of long distance (intercity route path) should pass the nodes of high-layer roads on the border. Thus the nodes near the border outside the test region can be selected as the representative nodes (e.g. F, T and B shown in Fig. 5). The reason that the nodes on the border are not selected is to keep the accelerating A* algorithm using the high-layer road network when the start point and destination locate in different middle-layer regions as the rule of the accelerating A* algorithm with vertical partitioning. Since some arterial roads may be missing in the high-layer road network, the proxy point of important cities around the test regions can be added to the point set to compensate the missing paths going through the nodes near the border. Moreover, it is tactically necessary to define a buffer zone encompassing the test region as shown in Fig. 6 to ensure all paths can be checked. The buffer zone can be composed of the neighboring regions in the first and the second order. For example, Tyrol State of Austria is a direct neighboring region of Bavaria State of Germany, while Trentino-Alto Adige of Italy is a neighboring region of the second order.

As shown in Fig. 5, different nodes are located in different regions and therefore bear different meanings. P represent a node on a middle-layer or high-layer road
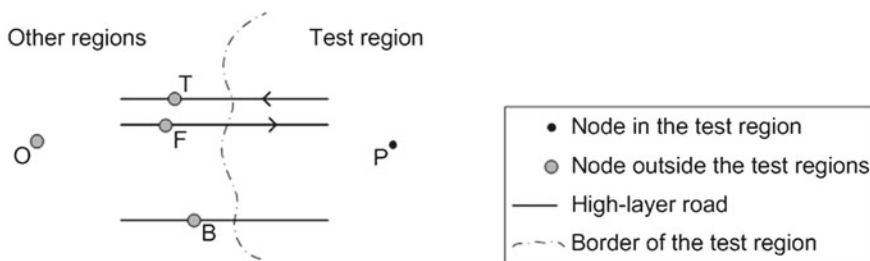
Fig. 5 Setting of starting points and destinations

Fig. 6 A schematic buffer zone encompassing the test area
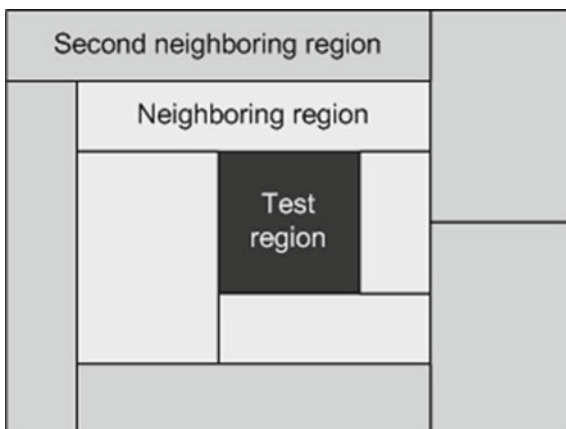
within or on the border of the test region; F represents a node on a dual carriageway near the border outside the test region with the direction towards the test region; T represents a node on a dual carriageways near the border outside the test region with the direction from the test region; B represents a node on a single carriageways near the border outside the test region; O represents a city in the buffer zone.

In order to keep the two-way access, all nodes in point sets should be used as both starting point and destination as the following sequence: (1) O -> P; (2) F -> P; (3) B -> P; (4) P -> O; (5) P -> T; (6) P -> B.

3.3 Decisive Path of OPCV Algorithm

The case demonstrated in Fig. 2 is in fact oversimplified. There are far more complex cases in the real road network. Considering the sketch in Fig. 8, the question how to pick up all roads in optimal paths that are missing in suboptimal paths is not trivial. Figure 8f shows the anticipated results of adjusted high-layer road network after conducting the OPCV test. Although the new high-layer road network is fit for the accelerating A* algorithm, the road network is broken
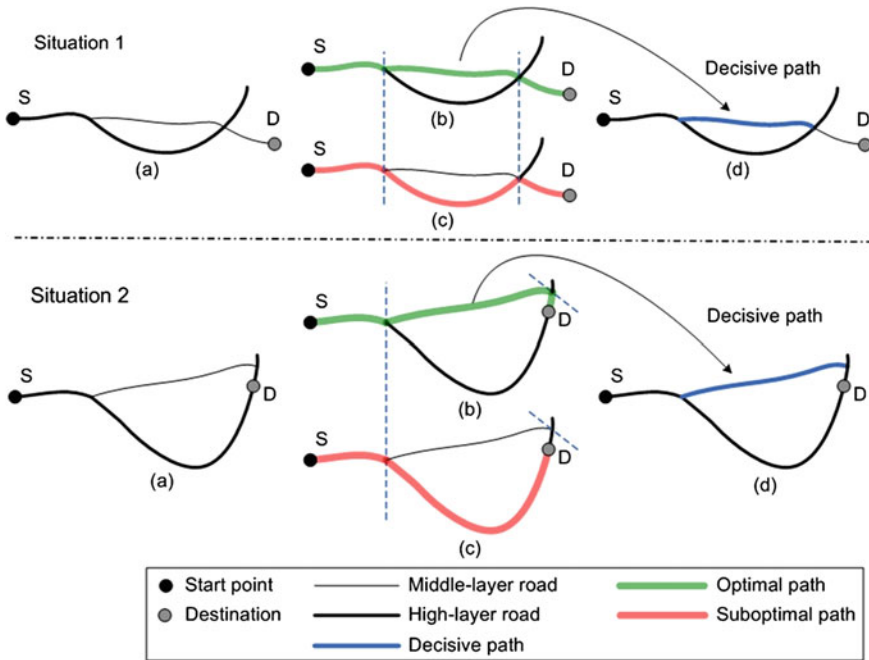
Fig. 7 Two situations of decisive path strategies. **a** Two-layer road network for improvement; **b** optimal path; **c** suboptimal path; **d** improved two-layer road networks

(e.g. Fig. 8h) and many extra roads would reduce the efficiency of route planning. To circumvent this drawback, we propose to identify a decisive path firstly. A decisive path exists in two situations:

1. Along the course of an optimal path, alternative roads that start and terminate at certain intermediate nodes of the optimal path may exist. The section of the optimal path between these nodes is termed as decisive path. Some examples are shown in Fig. 7 (Situation 1) and Fig. 8 (the route between node-0 and 1).
2. Different roads may exist from the same start point or terminate at the same destination when the start point or the destination of a route is a node on a high-layer road. In such cases, the middle-layer road constitutes the decisive path in the optimal path. Figure 7 (Situation 2) and Fig. 8 (the route between node-0 and 2) demonstrate two cases.

In addition, all decisive paths should start and terminate along the high-layer roads. Only in this way the connectivity with the original high-layer road network can be preserved. Once the decisive path is identified as shown in Fig. 8g, other optimal routes, e.g. from start node-0 to node-3, 4 and 5 would also be found based on the new high-layer road network. In another word, when a decisive path is picked up into high-layer road network, the related problems with suboptimal route are easily solved.
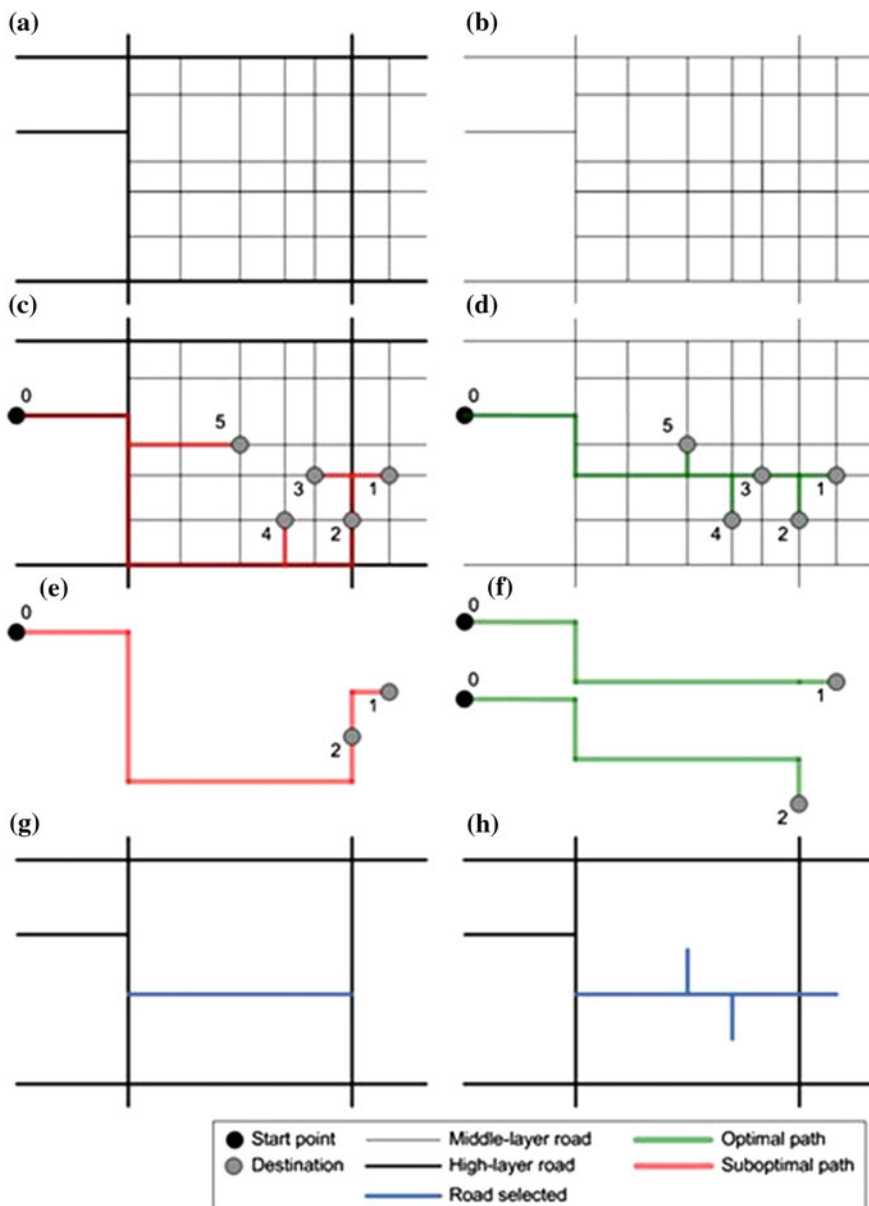
Fig. 8 Examples of selection by OPCV algorithm. **a** Two-layer road networks (*middle layer and high layer*); **b** single-layer road network (*middle layer*); **c, e** suboptimal paths calculated by accelerating A* algorithm; **d, f** optimal paths calculated by A* algorithm; **g** roads selected by decisive path; **h** roads selected by picking up all different roads in optimal paths

4 Case Study and Evaluation

4.1 Case Study in Bavaria State

The proposed approach is deployed for a test bed in Bavaria State in Germany, where the road dataset has a high quality with almost complete attribute values of road network. Therefore, we can easily get integrated higher-layer road networks with good connectivity by using the attribute "net 2 class" and compare the results with the existing reference road networks for verification. The test dataset is provided by Tele Atlas version 200610 and covers 10 countries including Germany, Czech, Austria, Liechtenstein, Switzerland, Netherlands, Belgium, Luxembourg, France and Italy.

In the initial step, the arterial roads of high-layer road network are extracted by connecting the important cities in the surrounding of the test area, as shown in Fig. 9, 13 cities are selected in surrounding regions of Bavaria State according to the freight capability and spatial distribution.

By applying the route-planning algorithm all arterial roads that go through these selected cities are identified as shown in Fig. 10a. The arterial roads within Bavaria State are highlighted in Fig. 10b. Apparently the road network with all arterial roads has a good legibility and allowable road density. As a bi-directional route planning between all cities is possible, every node in the road network can be
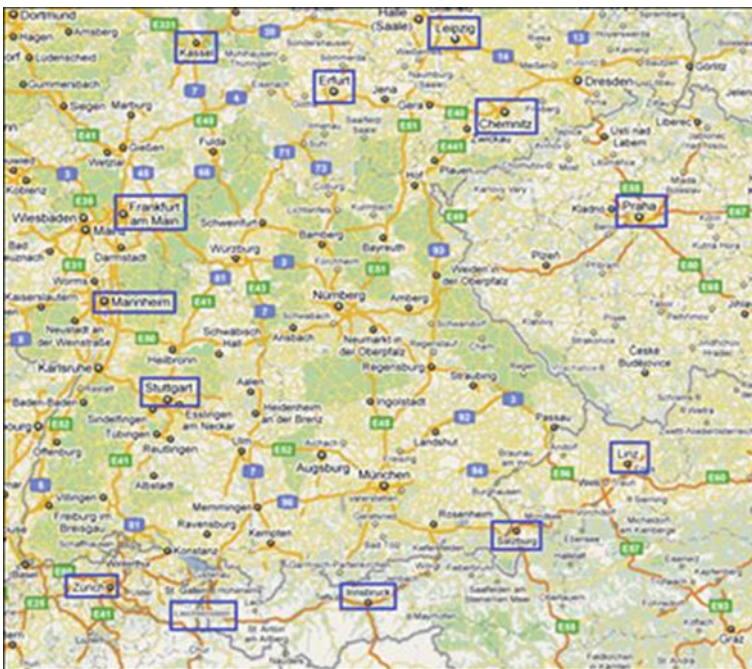
Fig. 9 Selected cities (in *blue frame*) outside of Bavaria state on a Google map
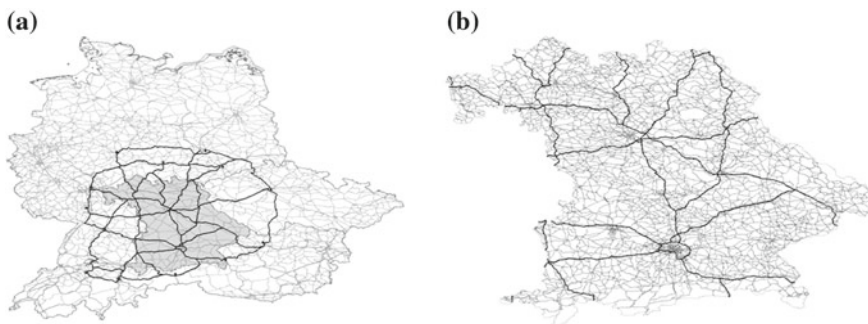
Fig. 10 Arterial roads in Bavaria state extracted by important cities. **a** Arterial roads (in *bold*) selected by important cities. **b** Arterial roads (in *bold*) selected by important cities in Bavaria state

reached from any other node. However, a few arterial roads may go lost due to the complexity of the road network or the lack of city selection, which however can be compensated by OPCV algorithm.

Figure 11 shows the result of detecting decisive paths by using OPCV algorithm. Totally, seven decisive paths are detected, which are ordered and highlighted in blue in Fig. 11. It should be note that the decisive paths (No.5 on Fig. 11) for the city Munich contain actually four pieces of decisive paths which can reflect the urban pattern of Munich.

Fig. 11 Decisive paths (in *blue*) detected by the OPCV algorithm in Bavaria state

(a) Result of OPCV algorithm.
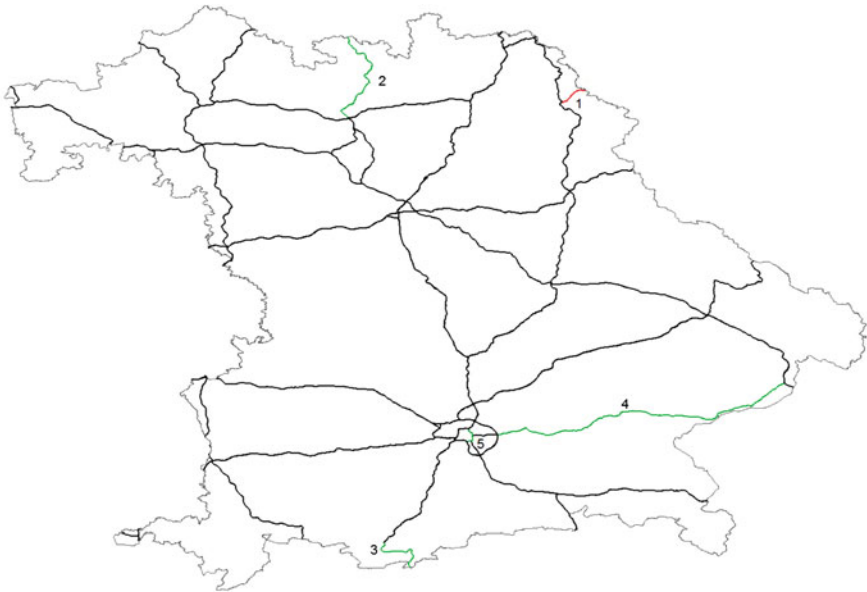
(b) Result of Navteq.

(c) Result of Tele Atlas.
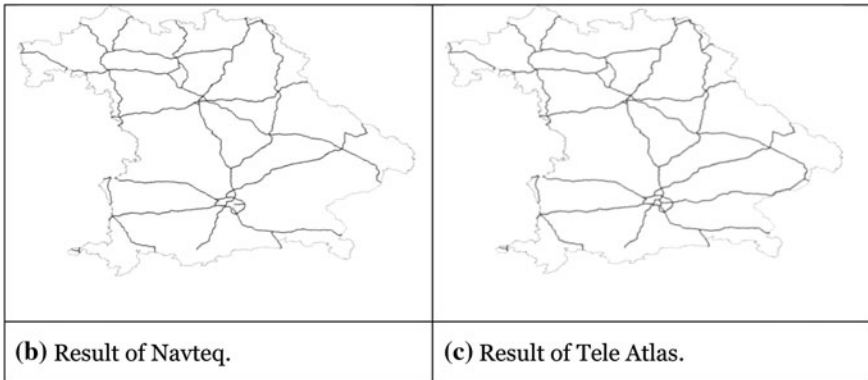
Fig. 12 Comparison of high-layer road networks in Bavaria state. **a** Result of OPCV algorithm. **b** Result of Navteq. **c** Result of Tele Atlas
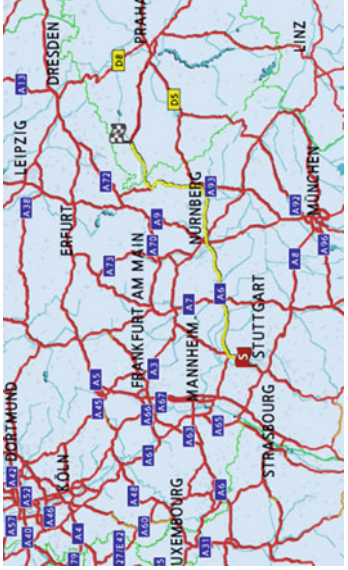
4.1.1 Evaluation

The results of our approach for test bed Bavaria State show a large similarity to those from Navteq's and Tele Atlas' which are tailored for navigation data authoring as shown in Fig. 12. Mainly, the differences exist for five routes, in which the route in red is not included in high-layer road network of our new reproach and the routes in green are contained.

On the base of the comparison above, four international routing paths (whose main parts are located in Bavaria State) are calculated using the proposed OPCV algorithm, as well as the algorithms of Navteq and Tele Atlas. In the following, these results are shown and analyzed with respect of their disagreements among the high-layer road networks. On the figures below, the calculated routes are visualized in yellow color with a start sign in red and an ending sign as checkered flag. In addition, the abbreviation TA stands for Tele Atlas and NA for Navteq.

- The route from Stuttgart Germany to Karlovy Vary in Czech

Route	Start point	Destination	Result	Difference
1	Stuttgart, Germany	Karlovy Vary, Czech	OPCV: OK TA:OK NA:OK	There is no difference among the results of the three algorithms, because they derived the same arterial road network along this route, as shown in Fig. 12

OPCV: OK

TA and NA: OK

- The route from Linz, Austria to Erfurt, Germany

Route	Start point	Destination	Result	Difference
2	Linz, Austria	Erfurt, Germany	OPCV: OK TA:NG NA:NG	The routes differ after Regensburg in Germany, because OPCV extracted a decisive path there (No. 2 on Fig. 12)

OPCV: OK

TA and NA: NG

- The route from Bolzano, Italy to North of Garmisch-Partenkirchen, Germany

Route	Start point	Destination	Result	Difference
3	Bolzano, Italy	North of Garmisch-Partenkirchen, Germany	OPCV: OK TA:NG NA:NG	The routes differ after Innsbruck, because OPCV extracts a decisive path there (No. 3 on Fig. 12)

OPCV: OK

TA and NA: NG

- The route from Vienna, Austria to Forstinning, Eastern Munich, Germany

Route	Start point	Destination	Result	Difference
4	Vienna, Austria	Forstinning, Eastern Munich, Germany	OPCV: OK TA:OK NA:NG	The routes differ from the boarder of German and Austria, because OPCV and TA extract a decisive path there (No. 4 on Fig. 12), while NA not

TA and OPCV: OK

NA: NG

5 Conclusion

Being inspired by the fact that the optimal performance of accelerating route-planning algorithms relies on whether the underlying road network is reasonably partitioned into vertical layers, an OPCV algorithm of selecting road network at the high-layer is proposed in this work.

The OPCV algorithm is a type of self-proving algorithm, so theoretically the road network of end result can support the corresponding accelerating route-planning algorithm for long-distance route. Technically, system tests can be triggered to confirm this characteristic.

Using the self-proving OPCV algorithm in a simulated test environment, sub-optimal routes found by the accelerating route-planning algorithms on an improperly partitioned road network can be automatically detected. Further, decisive paths that distinguish the suboptimal route from the optimal one can be identified and added to the high-layer network. The proposed approach is deployed for the test data in Bavaria State in Germany. The extracted road network at the high-layer reveals better connectivity of the arterial roads than those extracted by Tele Atlas and Navteq. This has been verified by calculating four routes going in or through Bavaria State.

References

Anders KH (2006) Grid typification. In: Riedl A, Kainz W, Elmes GA (eds) Progress in spatial data handling. 12th international symposium on spatial data handling (SDH), Vienna, 10–14 July

Chaudhry O, Mackaness WA (2005) Rural and urban road network generalization deriving 1:250,000 from 1:1250. In: International cartographic conference. Coruna, pp 9–16

Chen J, Hu YG, Li ZL, Zhao RL, Meng LQ (2009) Selective omission of road features based on mesh density for automatic map generalization. Int J Geogr Inf Sci 23(8):1013–1032

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1(1):269–271

Edwardes A, Mackaness W (2000) Modelling knowledge for automated generalization of categorical maps—a constraint based approach. In: Atkinson P, Martin D (eds) GIS and GeoComputation (innovations in GIS 7). Taylor & Francis, London

Gulgen F, Gokgoz T (2008) Selection of roads for cartographic generalization. In: The international archives of the photogrammetry, remote sensing and spatial information sciences, vol XXXVII, Part B4

Heinzle F, Anders KH (2007) Characterising space via pattern recognition techniques: identifying patterns in road networks. In: Ruas A, Sarjakoski T, Mackaness WA (eds) Generalisation of geographic information: cartographic modelling and applications. Elsevier Ltd, Oxford

Heinzle F, Sester M, Anders KH (2005) Graph-based approach for recognition of patterns and implicit information in road networks. In: Proceedings of 22nd international cartographic conference, La Coruña, pp 9–16

Heinzle F, Anders KH, Sester M (2006) Pattern recognition in road networks on the example of circular road detection. In: Proceedings of the 4th international conference GIScience, Münster

Hu YG, Chen J, Li ZL, Zhao RL (2007) Selection of streets based on mesh density for digital map generalization. In: Proceedings of the 4th international conference on image and graphics

Kulik L, Duckham M, Egenhofer M (2005) Ontology-driven map generalization. J Vis Lang Comput 16(3):245–267

Li ZL, Choi YH (2002) Topographic map generalization: association of road elimination with thematic attributes. Cartogr J 39(2):153–166

Liu XJ, Zhan BJ, Ai TH (2009) Road selection based on Voronoi diagrams and 'Strokes' in map generalization. Int J Appl Earth Obs Geoinf

Mackaness WA (1995) A constraint based approach to human computer interaction in automated cartography. In: Proceeding ICA/ACI, vol 2, Barcelona, pp 1423–1433

Mackaness W (2007) Understanding geographic space. In: Mackaness W, Raus A, Sarjakoski T (eds) The generalization of geographic information: models and applications. Elsevier, Amsterdam

Mackaness WA, Beard MK (1993) Use of graph theory to support map generalisation. Cartogr Geogr Inf Syst 20:210–221

Peng W, Muller JC (1996) A dynamic decision tree structure supporting urban road network automated generalisation. Cartogr J 33(1):5–10

Sester M (1995) Lernen struktureller Modelle für die Bildanalyse. Deutsche Geodätische Kommission, Reihe C, Nr. 441, München 1995, p 118

Sinha G, Flewelling D (2002) A framework for multicriteria line generalization to support scientific and engineering modeling. In: Egenhofer MJ, Mark DM (eds) GIScience 2002 abstracts, pp 173–175

Thom S (2005) A strategy for collapsing OS integrated transport network dual carriageways. In: Proceedings of the 8th ICA workshop on generalization and multiple representation, La Coruña

Thom S (2006) Conflict identification and representation for roads based on a skeleton. In: Proceeding of the 12th international symposium on spatial data handling 12, Vienna, pp 659–680

Thomson RC, Brooks R (2000) Efficient generalisation and abstraction of network data using perceptual grouping. In: Proceedings of the 5th international conference on GeoComputation, Chatham

Thomson R, Richardson D (1999) The "good continuation" principle of perceptual organization applied to the generalisation of road networks. In: Proceedings of the 19th ICC, ICA, Ottawa

Tian J (2008) Progressive representation and generalization of street network vector data. ISPRS Congress, Beijing

Touya G (2007) A road network selection process based on data enrichment and structure detection. in: Proceedings of the 10th ICA workshop on generalization and multiple representation, Moscow

Wertheimer M (1938). Laws of organization in perceptual forms. Untersuchungen zur Lehre von Der Gestalt II, in Psycologische Forschung 4:301–350

Zhang Q (2004) Modelling structure and patterns in road network generalization. In: proceedings of ICA workshop on generalisation and multiple representation, Leicester, 20–21 Aug 2004

Zhang X, Ai TH, Jantien S (2008) The evaluation of spatial distribution density in map generalization. ISPRS Congress, Beijing

A Propagating Update Method of Multi-Represented Vector Map Data Based on Spatial Objective Similarity and Unified Geographic Entity Code

Yanxia Wang, Qingyun Du, Fu Ren and Zhiyuan Zhao

Abstract In recent years the propagating update of multi-represented datasets has become a crucial issue for maintaining geographic data, especially since National Spatial Data Infrastructure (NSDI) appeared. The key to propagating update is building the mappings between the datasets. Usually the traditional approaches of building the mappings don't consider the attributive similarity and often use object IDs to build the mappings which may differ with the variation of data storage. Accordingly, a comprehensive similarity computing method is proposed and unified geographic entity code (UGEC) is put forwards to build the mappings in this chapter. A workflow of propagating update, which mainly consists of data preprocessing, changes detecting, changes extracting, master dataset updating, and target dataset updating, is presented on the basis of objects mappings. An experiment on implementation of this method demonstrates its viability at the end.

Keywords Propagating update · Similarity · Unified geographic entity code

1 Introduction

As National Spatial Data Infrastructure (NSDI) of many countries been successfully constructed in recent years, geospatial data has been applied in all aspects of global businesses. Accordingly the core of GIS has made a transition from data

Y. Wang · Q. Du
School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

F. Ren (✉)
School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China
e-mail: renfu@whu.edu.cn

Z. Zhao
State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China

production to data update and data update is an important factor of sustainable development of GIS (Fritsch 1999).

NSDI usually contains different types and scales of spatial datasets, in which the same entity is represented in multiple ways such as different geometries, scales, semantics and spatial relations (Volz 2006). Traditionally the multi-represented datasets are updated according to the scale or type respectively, which makes the inconsistency of multi-represented datasets and is a labor-intensive, time- and money-consuming procedure (Qi et al. 2010). In order to mitigate these problems, propagating update is proposed, i.e. when a dataset is updated, the corresponding update information can be propagated to its correlated datasets.

Propagating update was firstly proposed by Harrie and Hellström (1999a, b) and a prototype system of four steps was introduced: examination, propagation, generalization and conflicts solution. Haunert and Sester (2005) developed this prototype system and showed an incremental approach of propagating update from a topographic source dataset to a generalized dataset. Links were applied to express the features' correspondences of the two datasets. Further Qi et al. (2010) presented a propagating update method from larger-scale map to smaller-scale map based on automatic changes detecting. Kang et al. (2004) discussed a new propagating update mechanism of multi-scale databases covering the same geographic area with different scales and being derived from an original one.

All of the aforementioned propagating update methods neglect the automatic method of building links between the multi-scale datasets. Many researchers discussed the similarity in feature matching (Al-Bakri and Fairbairn 2012; Belussi et al. 2005; Frontiera et al. 2008; Walter and Fritsch 1999) to build the mappings. Tversky (1977) introduced two similarity computing models—contrast model and ratio model, which had made a great impact on latter similarity studies. Samal et al. (2004) brought geographic context similarity to similarity computing. Li and Fonseca (2006) built the TDD (topology, direction, distance) model to assess similarity. Li and Goodchild (2011) developed a new optimization model to improve the linear matching by extending the optimized feature matching method which had been proposed in 2010 (Li and Goodchild 2010), emphasizing the high percentage of correct matching than greedy matching.

Because of the features matched method based on similarity, similarity was brought in propagating update. Sheeren et al. (2009) pointed out that data update was one of the important applied directions of geometric feature matching. Wang and Wei (2008) computed the spatial similarity of instances from the MDB (Master Database) and CDB (Client Database) to match the schemas automatically. Further update information was propagated from MDB to CDB depending on the schema mappings.

Accordingly, this chapter introduces a propagating update method of multi-represented vector data on the basis of spatial objective similarity and Unified Geographic Entity Code (UGEC).

2 Mappings of Objects

The key of propagating update is to build the mappings between multi-represented datasets. Lamentably the mappings were not recorded as the databases building. The mappings are built manually through the links of IDs traditionally. But the IDs are automatically produced by computer and may differ with the variation of data storage. Hence the mappings are not steady. A new method of building mappings should be proposed to solve the problem.

2.1 Spatial Objective Similarity

Spatial similarity is that the regions are considered similar at a particular granularity (scale) and context (thematic properties) (Holt 1999; Holt and Benwell 1997). This is compared from scene. But the similarity can also be compared from object, which is called spatial objective similarity. To illustrate it, set theory is applied to define: suppose there are two objects A and B. The characteristic sets of A and B are C_A and C_B, where $C_A \neq \emptyset$ and $C_B \neq \emptyset$. If $C = C_A \cap C_B \neq \emptyset$, then the characteristic set C is the similar relation of A and B.

Spatial objective similarity contains geometric similarity, attributive similarity and context similarity. The context similarity is the similar distances and directional relationships with other objects in their respective datasets (Samal et al. 2004). It is complex, whereas the geometric similarity and the attributive similarity are enough to identify the matching objects. Accordingly the geometric similarity and the attributive similarity are chosen to be the factors of computing spatial objective similarity. The similarity formula from object B to object A is as follows:

$$S(B,A) = W_{\mathrm{geo}} * S_{\mathrm{geo}}(B,A) + W_{\mathrm{att}} * S_{\mathrm{att}}(B,A)$$

where $S_{\mathrm{geo}}(B,A)$ is the geometric similarity and $S_{\mathrm{att}}(B,A)$ is the attributive similarity. W_{geo} and W_{att} are the weight of $S_{\mathrm{geo}}(B,A)$ and $S_{\mathrm{att}}(B,A)$ respectively. And $S(B,A), S_{\mathrm{geo}}(B,A), S_{\mathrm{att}}(B,A) \in [0,1], W_{\mathrm{geo}} + W_{\mathrm{att}} = 1$.

1. Geometric similarity $S_{\mathrm{geo}}(B, A)$

   The factors of geometric similarity are centroid position, area, Minimum Bounding Rectangle (MBR), angle and so on. The geometric similarity formula is related to the geometry type as follows.

   - Between area object A and area object B (Fu and Wu 2008)

$$S_{geo}(B,A) = \alpha \times \frac{S(A \cap B)}{S(A)} + \beta \times \frac{\min[S(A), S(B)]}{\max[S(A), S(B)]} + \gamma \times \left(1 - \frac{d}{l}\right)$$

where S means computing the area, d means the distance between the centroids of A and B, l means the diagonal length of A's MBR, and $\alpha + \beta + \gamma = 1$.

- Between area object A and line object B:

$$S_{geo}(B,A) = \alpha \times \frac{\min\{L[cl(A)], L(B)\}}{\max\{L[cl(A)], L(B)\}} + \beta \times \left\{1 - \frac{|\theta[cl(A)] - \theta(B)|}{|\theta[cl(A)]|}\right\} + \gamma$$
$$\times \left(1 - \frac{d}{l}\right)$$

where L means computing the length, cl means the central axis, θ means computing the direction, d means the distance between the centroids of A and B, l means the diagonal length of A's MBR, and $\alpha + \beta + \gamma = 1$.

- Between surface object A and point object B:

$$S_{geo}(B,A) = 1 - \frac{d}{l}$$

where d means the distance between A's centroid and B, l means the diagonal length of A's MBR.

- Between line object A and line object B (Li and Fonseca 2006; Fu and Wu 2008):

$$S_{geo}(B,A) = \alpha \times \left[1 - \frac{|L(B) - L(A)|}{L(A)}\right] + \beta \times \frac{S(Buf(A) \cap Buf(B))}{S(Buf(A))} + \gamma$$
$$\times \left[1 - \frac{|\theta(B) - \theta(A)|}{|\theta(A)|}\right]$$

where L means computing the length, Buf means the buffer zone, S means computing the area, θ means computing the direction, and $\alpha + \beta + \gamma = 1$.

- Between line object A and point object B:

$$S_{geo}(B,A) = 1 - \frac{d}{L(A)}$$

where d means the distance between A's centroid and B, L means computing the length.

- Between point object A and point object B (Samal et al. 2004):

$$S_{geo}(B,A) = 1 - \frac{\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}}{U}$$

where (x_A, y_A), (x_B, y_B) are respectively the coordinates of A and B, U is a normalization factor.

2. Attributive similarity $S_{att}(B, A)$

Attributive similarity describes the proximity of attribute values between A and B. Suppose X is the shared attribute set, and the values of A and B are

x_A, x_B respectively. Generally, the types of attribute value may be numeral, enumeration or character. The different types of attribute value will have different computing formulas:

- If X is numeral, and suppose $x_B \geq x_A$ (Li et al. 2005):

$$S_{att}(B,A) = 1 - \frac{\int_{x_A}^{x_B} \sqrt{\left(\left[f(x)'\right]^2 + 1\right)} dx}{\int_{x_{\min}}^{x_{\max}} \sqrt{\left(\left[f(x)'\right]^2 + 1\right)} dx}$$

where $f(x)$ is the distribution function of attribute X in the interval of $[x_{min}, x_{max}]$.

- If X is enumeration:

$$S_{att}(B,A) = 1 - \frac{|Seq(x_B) - Seq(x_A)|}{N}$$

where Seq means computing the sequence number in the enumeration set, and N is the number of enumeration.

- If X is character:

$$S_{att}(B,A) = \begin{cases} 1, & if\ string(x_A) = string(x_B) \\ 0, & if\ string(x_A) \neq string(x_B) \end{cases}$$

For simplicity in computing, the attribute similarity is computed by judging whether they are identical.

2.2 Unified Geographic Entity Code

In order to identify the geographic entity uniquely and uniformly, a unified geographic entity encoding method is put forward. The code is called as unified geographic entity code (UGEC), which consists of geographic entity unique code, surveying industry code, and represented scale code.

- Geographic entity unique code includes geographic grid number, centroid coordinate, the length and width of MBR.
- Surveying industry code consults the geographic entity code of the surveying and mapping industry standard—"Platform for geo-information common services Data specification for geo-entity, geographic name and address".
- Represented scale is the code of scale in which the geographic entity should be represented in map.

It can identify any entity uniquely and recognize the same entity in different datasets uniformly.

2.3 Building the Mappings

The mappings between the multi-represented datasets should be built before making the updates propagating. There are two main methods: (1) manually building the mappings through visual interpretation; (2) (semi-)automatically building the mappings through objects matching (Kieler et al. 2009; Rainsford and Mackaness 2002).

A new item of UGEC is added in all datasets. Firstly, the value of the most detailed dataset's adding item is manually entered according the encoding regulation (Sect. 2.2). Then the object's buffer set is built to define the candidate matching objects in less detailed datasets and the similarities of candidate matching pairs are computed (Sect. 2.1). The matching pair is found by comparing the similarity with threshold. The UGEC of the object in the less detailed datasets is entered automatically according the matching object's UGEC and encoding regulation. The similarities and differences of attribute structures between the multi-represented datasets are then analyzed and the mapping table is established. Therefore the mappings between multi-represented datasets are built.

3 Propagating Update

Propagating update is that the update information is propagated among the multi-represented datasets. It firstly updates the more detailed dataset, and then updates the less detailed or the same detailed dataset according the mappings. In order to describe clearly, the dataset updated in advance is defined as "master dataset", and the propagating updated dataset as "target dataset".

The propagating update workflow (Fig. 1) is as follows:

1. Preprocess data
   Firstly the current master dataset is extracted from the current database. And the multi-represented datasets is preprocessed, such as making the spatial references coordinate systems between the up-to-date dataset and the current master dataset consistency, building the mapping table of attribute structures, filtering the useless features for the current master dataset in the up-to-date dataset and so on.
2. Detect changes
   In this step, the spatial objective similarity between up-to-date dataset and current master dataset is computed, and the changes are detected (suppose O_N is an object in up-to-date dataset):

   - If there is not any matching object in current master dataset, O_N is a new addition.
   - If there is a matching object in current master dataset, and supposed that O_M is the matching object. Then it judges that whether the similarity $S(O_M, O_N)$ is
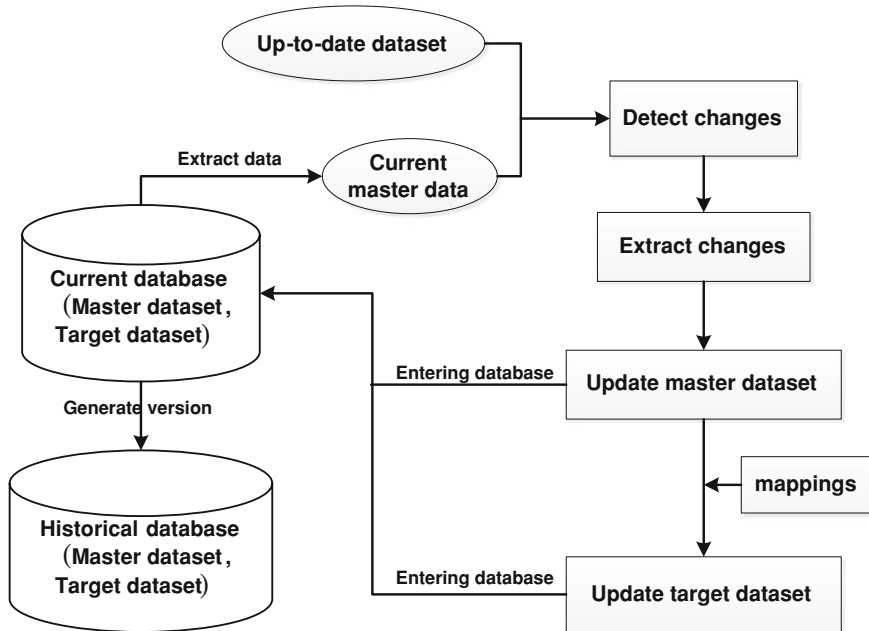
Fig. 1 Workflow of propagating update

larger than the threshold T_1(e.g. 0.9).(1) If $S(O_M, O_N) \geq T_1$, then O_M is invariable. (2) If $0 < S(O_M, O_N) < T_1$, then O_M is an alteration. And if $S_{geo}(O_M, O_N) \geq T_1$, only attribute is variable. If $S_{att}(O_M, O_N) \geq T_1$, only geometry is variable. Otherwise the geometry and the attribute are both variable.

- If there is no object in up-to-date dataset matching O_M, that is for any O_N, $S(O_M, O_N) = 0$, then O_M is a disappearance.

3. Extract changes

According the judgment of similarity, the changed objects' (new additions, alterations and disappearances) information is extracted.

4. Update master dataset

Variable *flag* is applied to mark the update operation, and its default value is 0. The operation of update master dataset is as follows.

- If O_M is a disappearance, then delete its information but the UGEC, and assign 1 to *flag*.
- If O_N is a new addition, then add an object with the extracting information, and assign 2 to *flag*.
- If O_N is an alteration, moreover it only has variation in geometry, then change the geometric information of O_M according the extraction, and assign 3 to

flag. Moreover, it only has variation in attribute, then change the attributive information of O_M according the extraction, and assign 4 to *flag*. Moreover it has variation both in geometric information and attributive information, then delete the old information, and add the new information according the extraction, and assign 5 to *flag*.

After updating, the consistency of the master dataset should be examined to eliminate the conflicts.

5. Update target dataset

   After completing the master dataset update (Fig. 2a), the update information will be propagated to the target dataset (Fig. 2b) according the update tags (*flag*) in master dataset and the mappings between the master dataset and the target dataset. Suppose O_M is also an object in the master dataset, and O_T is the object in target dataset matching O_M, and T_2 (e.g. 0.85) is the threshold of similarity.

   - If the *flag* of O_M is 1, then delete the information of O_T, but persist its UGEC. And assign 1 to *flag*. For example, the object 4 (Fig. 2b) is a disappearance. It should be deleted in the target dataset (Fig. 2c).
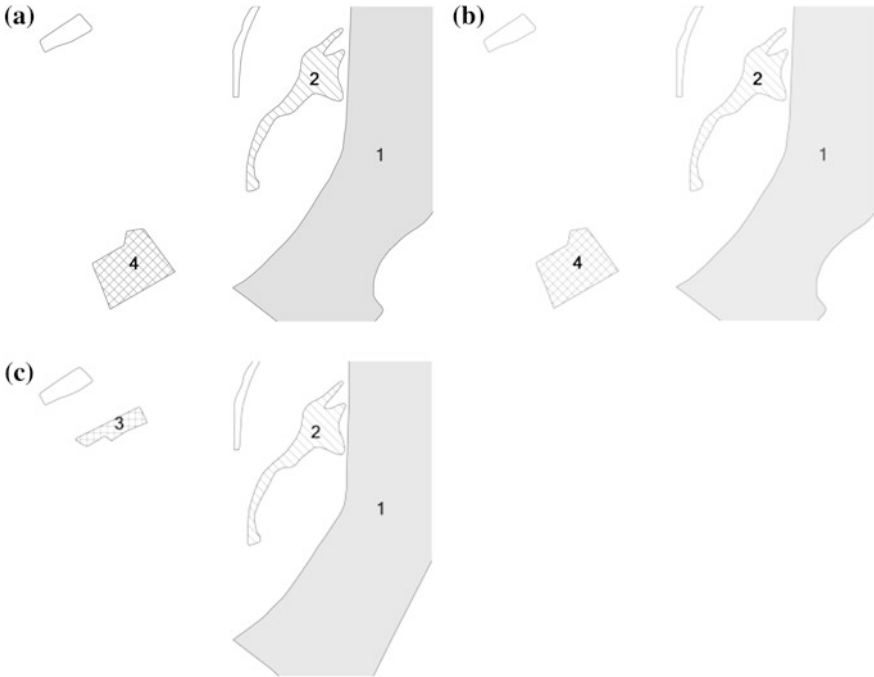
Fig. 2 The sample of updating target dataset. **a** Updated master dataset. **b** Old target dataset. **c** Updated target dataset

- If the *flag* of O_M is 2, then extract the information of O_M, process the geometric information such as generalization, following add the object to the target dataset with its processed information, at the same time add its UGEC. At last the *flag* of O_T is set as 2. For example, object 3 (Fig. 2a) is a new addition. It should be generalized and added in the target dataset (Fig. 2c).
- If the *flag* of O_M is 3, then compute the geometric similarity $S_{geo}(O_T, O_M)$ between O_T and O_M. If $S_{geo}(O_T, O_M) \geq T_2$, the change may be not displayed in the target dataset, so it should not be updated to the target dataset, and the *flag* of O_T is set as 0. If $S_{geo}(O_T, O_M) < T_2$, extract the geometric information of O_M, process the geometric information, then update the object O_T in the target dataset with the processed information. And the *flag* of O_T is set as 3. For example, object 2 is a geometric alteration (Fig. 2a), and $S_{geo}(O_{T2}, O_{M2}) \geq T_2$, so it should not be changed in target dataset (Fig. 2c). But object 1 is also a geometric alteration, and $S_{geo}(O_{T2}, O_{M2}) < T_2$, so it should be updated in target dataset (Fig. 2c).
- If the *flag* of O_M is 4, then extract the attributive information of O_M to update O_T in the target dataset. Further assign 4 to the *flag* of O_T.
- If the flag of O_M is 5, then compute the geometric similarity $S_{geo}(O_T, O_M)$ between O_T and O_M. If $S_{geo}(O_T, O_M) \geq T_2$, then extract the attributive information of O_M to update O_T, and set the *flag* of O_T as 4. If $S_{geo}(O_T, O_M) < T_2$, then delete the geometric information and variable attributes. It extracts the geometric information and attributive information of O_M, and process the information. Then it updates O_T with the processed information and set *flag* as 5. After propagating update information to the target dataset, it should examine the consistency to eliminate the conflicts.

    It can update the other target datasets by analogy. After all target datasets updates are finished, all *flag*s will be set as 0.

6. Enter database

    At the end, it should enter all the updated data to current database to form new "master dataset" and "target dataset". If it is needed to form a new version dataset, it should also generate new version dataset and store it in the historical database, and set up a snapshot rollback mechanism to manage the data versions. Therefore, the data is updated persistently, steadily, and safely.

4 Experiment

Three datasets applied as test data are as follows:

1. Building census dataset: it is about the buildings' information, which consists of location, geometry, name, property, floor space, and so on (Fig. 3c).
2. 1:2000 electronic map: it is the electronic map in the scale of 1:2000, whose features include hydrographic network, settlement lands and facilities,

Fig. 3 Propagating update test datasets. **a** 1:2000 electronic map. **b** 1:5000 electronic map. **c** Building census dataset

transportation, line pipe, administrative boundary and region, landform, vegetation and land property, POI (point of interest) etc. (Fig. 3a).

3. 1:5000 electronic map: it is the electronic map in the scale of 1:5000 which has the same features as the 1:2000 electronic map (Fig. 3b).

The same building may be represented in these three datasets. Therefore the buildings update information can be propagated from building census dataset to 1:2000 electronic map and 1:5000 electronic map. Supposed 1:2000 and 1:5000 electronic maps have the same up-to-dateness (Fig. 3a, b). And building census dataset (Fig. 3c) is newer than the electronic maps.

A prototype system is built to realize propagating update between multi-represented vector datasets. Firstly the similarity threshold T_0 is set as 0.85 to build the mappings of 1:2000 electronic map and 1:5000 electronic map (Chap. 2). It sets the similarity threshold T_1 as 0.9 to update 1:2000 electronic map (the feature of settlement lands and facilities) through the prototype system according spatial objective similarity (Fig. 4). Further the *flag*s of updated objects are set at the same. Then the similarity threshold T_2 is set as 0.85, and the changes are

Fig. 4 Updated 1:2000 electronic map (updated objects with *red* boundaries)
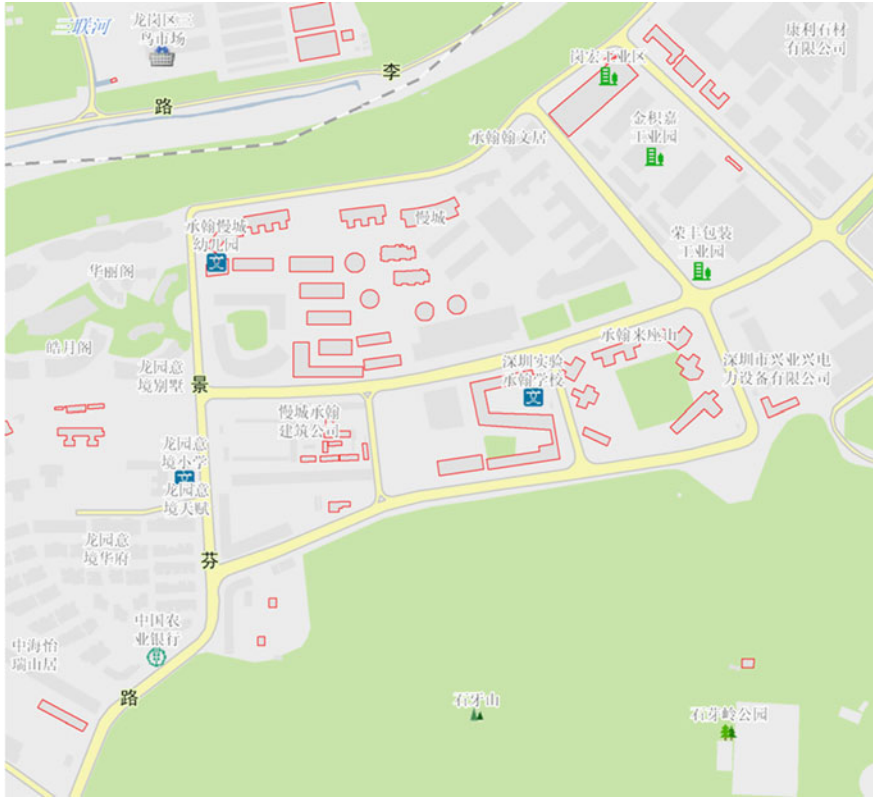
Fig. 5 Updated 1:5000 electronic map (updated objects with *red* boundaries)

propagated to 1:5000 electronic map (the feature of settlement lands and facilities) by this prototype system according *flags*, object mappings and spatial objective similarity (Fig. 5).

There are also two datasets of 1:2000 electronic map and 1:5000 electronic map updated manually. The automatically updated electronic maps superpose the manually updated electronic maps according the scale respectively (Fig. 6). From the overlay of these electronic maps, it can be seen that the settlement lands and facilities features of 1:2000 maps are nearly unanimous, but 1:5000 maps have a few inconsistencies because of automatic generalization. Hence this propagating updating method is verified effective, accurate and viable.

Fig. 6 Overlay of electronic maps (updated objects with *red* boundaries, invariable objects with *blue* boundaries, the inconsistencies of overlay filling with *black*). **a** Overlay of 1:2000 electronic maps. **b** Overlay of 1:5000 electronic maps

5 Conclusion

Nowadays, many countries have completed the building of geo-spatial database. And the up-to-dateness directly affects the value of data and indirectly affects the value of geographic information system. Along with the construction of NSDI, data become more and more "intensive" and shared. Meanwhile the independent updating between multi-represented datasets causes many problems. Therefore the multi-represented data propagating update mechanism is essential to the development of NSDI.

This chapter proposes a comprehensive computing method for spatial objective similarity and UGEC to build the mappings between vector datasets, through which propagating updates between multi-represented datasets can be realized. This method is proved to be helpful to the improvement of the quality and efficiency of multi-represented data updating work in NSDI data maintenance.

References

Al-Bakri M, Fairbairn D (2012) Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. Int J Geogr Inf Sci 26(8):1–20

Belussi A, Catania B, Podesta P (2005) Towards topological consistency and similarity of multiresolution geographical maps. In: Proceedings of the 13th annual ACM international workshop on Geographic information systems, pp 220–229

Fritsch D (1999) GIS data revision-visions and reality: keynote speech in joint ISPRS commission workshop on dynamic and multi-dimensional GIS. NGCC, Bejing

Frontiera P, Larson R, Radke J (2008) A comparison of geometric approaches to assessing spatial similarity for GIR. Int J Geogr Inf Sci 22(3):337–360

Fu ZL, Wu JH (2008) Entity matching in vector spatial data. Vol. XXXVII. Part B4, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Bejing, pp 1467–1472

Harrie L, Hellström AK (1999a) A case study of propagating updates between cartographic data sets. In: Proceedings of the 19th International Cartographic Conference of the ICA, Ottawa, Canada

Harrie L, Hellström AK (1999b) A prototype system for propagating updates between cartographic data sets. Cartogr J 36(2):133–140

Haunert JH, Sester M (2005) Propagating updates between linked datasets of different scales. In: Proceedings of XXII international cartographic conference (ICC 2005), pp 11–16

Holt A (1999) Spatial similarity and GIS: the grouping of spatial kinds. In: 11th annual colloquium of the spatial information research center (SIRC05), pp 241–250

Holt A, Benwell GL (1997) Using spatial similarity for exploratory spatial data analysis: some directions. In: Proceedings of the 13th annual ACM international workshop on geographic information systems, pp 279–288

Kang HK, Moon JW, Li KJ (2004) Data update across multi-scale databases. In: Proceedings of the 12th international conference on geoinformatics, pp 749–756

Kieler B, Huang W, Haunert JH, Jiang J (2009) Matching river datasets of different scales. In: Sester M, Bernard L, Paelke V (eds) Advances in GIScience. Springer, Berlin, pp 135–154

Li B, Fonseca F (2006) Tdd: a comprehensive model for qualitative spatial similarity assessment. Spatial Cogn Comput 6(1):31–62

Li L, Goodchild MF (2010) Automatically and accurately matching objects in geospatial datasets. In: Proceedings of theory, data handling and modelling in geospatial information science, Hong Kong, pp 26–28

Li L, Goodchild MF (2011) An optimisation model for linear feature matching in geographical data conflation. Int J Image Data Fusion 2(4):309–328

Li F, Zhou KB, Feng S (2005) A similarity calculation strategy based on the statistic of case feature. J Huazhong Univ Sci Tech (Nat Sci Ed) 33(6):80–82

Qi HB, Li ZL, Chen J (2010) Automated change detection for updating settlements at smaller-scale maps from updated larger-scale maps. J Spatial Sci 55(1):133–146

Rainsford D, Mackaness W (2002) Template matching in support of generalisation of rural buildings. In: Advances in spatial data handling. In: Proceedings 10th international symposium on spatial data handling, pp 137–151

Samal A, Seth S, Cueto K (2004) A feature-based approach to conflation of geospatial sources. Int J Geogr Inf Sci 18(5):459–489

Sheeren D, Mustière S, Zucker JD (2009) A data-mining approach for assessing consistency between multiple representations in spatial databases. Int J Geogr Inf Sci 23(8):961–992

Tversky A (1977) Features of similarity. Psychol Rev 84(4):327–352

Volz S (2006) An iterative approach for matching multiple representations of street data, pp 101–110

Walter V, Fritsch D (1999) Matching spatial data sets: a statistical approach. Int J Geogr Inf Sci 5(13):445–473

Wang YH, Wei FY (2008) A schema-matching-based approach to propagating updates between heterogeneous spatial databases 714605-1-10

Part IV
Higher Dimensional Visualisation and Augmented Reality

Visualization of Trajectory Attributes in Space–Time Cube and Trajectory Wall

Gennady Andrienko, Natalia Andrienko, Heidrun Schumann and Christian Tominski

Abstract Space–time cube is often used as a visualization technique representing trajectories of moving objects in (geographic) space and time by three display dimensions (Hägerstrand 1970). Despite the recent advances allowing space–time cube visualization of clusters of trajectories, it is problematic to represent trajectory attributes. We propose a new time transformation—sequential ordering—that transforms the space–time cube into a new display, trajectory wall, which allows effective and efficient visualization of trajectory attributes for trajectories following similar routes. To enable temporal analysis regarding temporal cycles, we use a time lens technique for interactive visualization. We demonstrate the work of the method on a real data set with trajectories of cars in a big city.

Keywords Movement data · Trajectories · Space–time cube · Trajectory wall

1 Introduction

Interactive space–time cube (STC) has become a common technique for visualizing trajectories (Kraak 2003; Kapler and Wright 2005; Andrienko et al. 2003). STC can support comparison of spatial, temporal, and dynamic properties (e.g., speed variation) of several trajectories when they are close in time. However, exploration of a large number of trajectories distributed over a long time period is challenging. Recent papers (Andrienko and Andrienko 2010, 2011) propose to use space–time cube in combination with trajectory clustering by route similarity

G. Andrienko (✉) · N. Andrienko
Fraunhofer IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
e-mail: gennady.andrienko@iais.fraunhofer.de

H. Schumann · C. Tominski
University of Rostock, Albert-Einstein-Straße 22, 18059 Rostock, Germany

(Rinzivillo et al. 2008) and transformation of absolute time references to relative positions within a temporal cycle (daily, weekly, yearly, etc.) or within the individual lifelines of the trajectories (e.g., temporal distances from the start time).

In this chapter, we suggest another visualization technique that can be viewed as extending the space–time cube by applying a particular time transformation. The main idea is to represent a set of spatially similar trajectories as a stack of bands, which resembles a wall, giving the name to the technique: trajectory wall (Tominski et al. 2012). As in the STC, two display dimensions represent two spatial dimensions. The third display dimension is divided into bins so that each bin contains one band representing one trajectory. The bands are divided into segments, which are coloured according to values of thematic attributes related to trajectory positions, such as speed, acceleration, or distance to the nearest neighbour (henceforth called positional attributes). The trajectories can be ordered in the third dimension according to their temporal order. Hence, a trajectory wall can be viewed as a space–time cube where the absolute time is transformed to the temporal order of the trajectories.

The trajectory wall technique is effective for spatially similar trajectories. Groups of spatially similar trajectories can be obtained by means of clustering (Andrienko et al. 2007; Rinzivillo et al. 2008; Schreck et al. 2008).

To represent values of a numeric positional attribute by colours of band segments, the attribute value range is divided into intervals, or classes. Several methods for defining classes are used in cartography (Slocum et al. 2009), including domain-specific class breaks, natural breaks, equal intervals, and equal class sizes. Andrienko and Andrienko (2006) proposed to use cumulative frequency curves for defining intervals interactively so as to obtain classes with similar cumulative measures computed by summing values of selected quantitative attributes. For trajectories, suitable cumulative measures are, for instance, total duration or total travelled distance.

2 Approach

For introducing our approach, we use a real data set of car traffic in Milan, Italy, during one day. We cluster the car trajectories according to route similarity and select one cluster consisting of 118 trajectories of the cars that travelled on the belt road counter-clockwise. The speed values for these trajectories are shown in a space–time cube in Fig. 1. The colours have been assigned to the class intervals <5 km/h, 5–10, 10–15, 15–30, 30–50, 50–75, 75–100, >100 km/h according to one of the Color Brewer colour scales (Harrower and Brewer 2003).

Figure 1 shows that the speeds within the given cluster are usually rather high, with some occasional exceptions during the day appearing as spots of yellow and red colours. It can be noticed that many trajectories had rather low speeds in the same area in close times. This indicates a traffic jam.
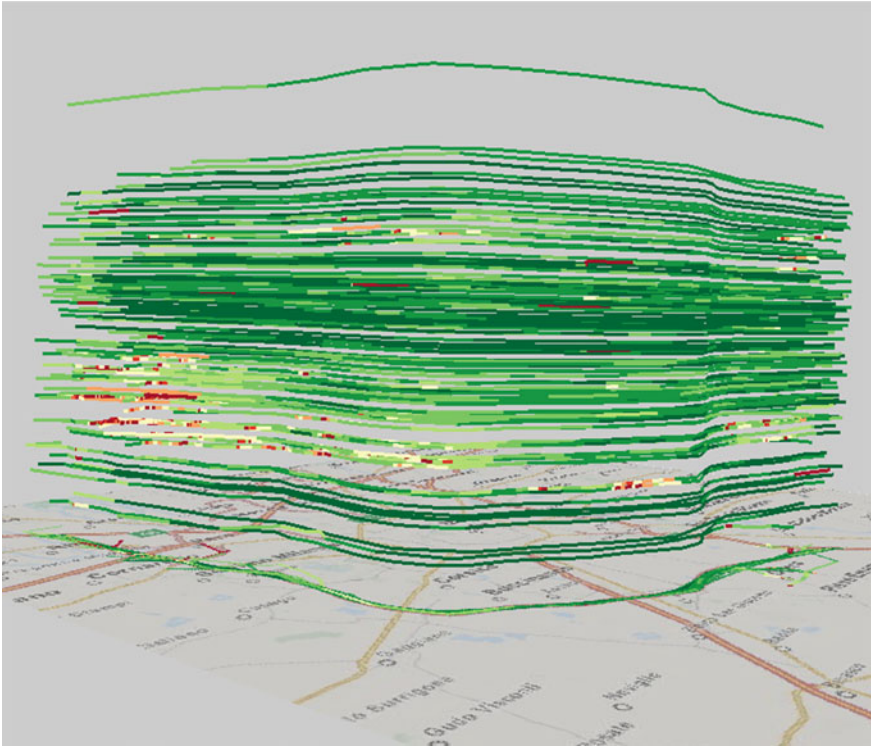
Fig. 1 Space–time cube shows the speed dynamics for a cluster of trajectories passing the city along the belt road counter-clockwise (from *left* to *right*). The colour legend is provided in Fig. 3

The space–time cube representation has several disadvantages. First, the lines may overlap, and therefore it is impossible to quantify the amount of trajectories affected by the traffic jam. Second, the line may intersect (as cars have different speeds, see Fig. 2), therefore it is hard to trace the speeds along trajectories.

It appears logical to eliminate occlusions and intersections of lines in the space–time cube by using the stacking layout. For representing two-dimensional geographic space, we need two display dimensions. We can add one more display dimension and use it for stacking visual elements representing trajectories. These may be segmented bands such that their shapes and positions with respect to the two spatial dimensions of the display correspond to the spatial properties of the trajectories. Each trajectory receives its individual portion of the third display dimension. The bands representing different trajectories are stacked one upon another; hence, the bands of different trajectories do not overlap. An example is shown in Fig. 3.

The display is oriented so that the northwest is on the left and the southeast on the right. The bands are ordered from bottom to top according to the start times of the respective trajectories. The colouring of the band segments encodes the values
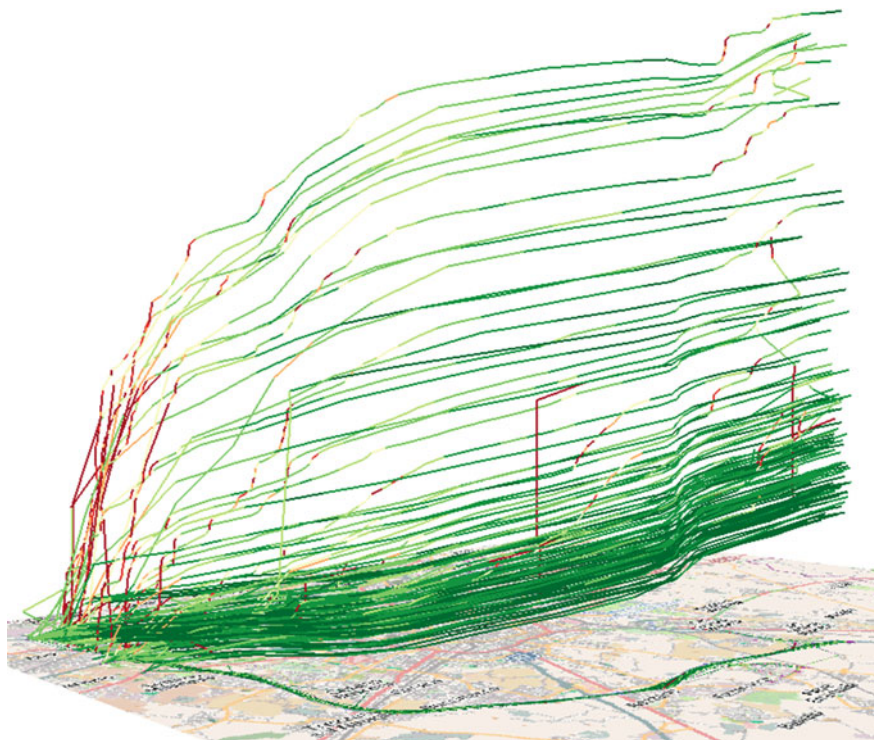
Fig. 2 The same trajectories as in Fig. 1 are shown with starting times aligned. A bunch of trajectories involved in traffic jam is easy to identify. The colour legend is provided in Fig. 3

of the positional attribute "speed". High speed values are represented by shades of green and low values by orange and red; yellow corresponds to the speeds between 15 and 30 km/h. The bands where all or almost all segments have the same colour (green) represent trajectories with uniform high-speed movement. The intrusions of yellow, orange, and red colours indicate that the movement slowed down.

We have not only an elementary view of the speeds in each individual trajectory but also an overall view of the distribution of the speeds over the space and across the multiple trajectories. We see the places where many cars slowed down. Low speeds mostly occur in neighbouring trajectories in the stack. Since the trajectories are ordered according to their start times, the vertical dimension of the display partly conveys the temporal component of the data. Hence, closeness of segments representing low speed values in the display may mean that these values are clustered in both space and time. The big spot of reduced speeds in the lower left part of the display may signify a prolonged traffic jam in the morning. An optional information display overlaid on the trajectory wall can show detailed information about the trajectory and segment currently pointed with the mouse. The information includes the temporal references, which helps us to locate traffic jams in time. We find that, indeed, the congestion on the northwest happened in the morning from about 5:40
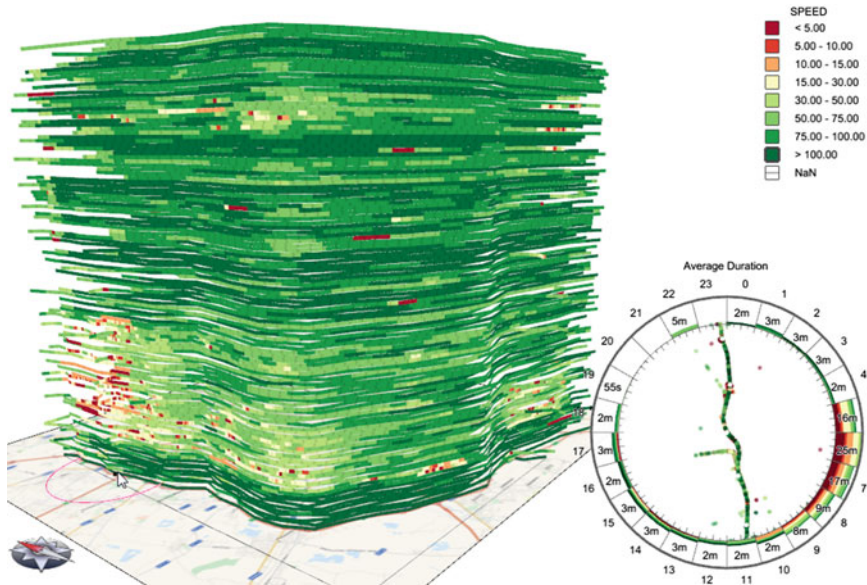
Fig. 3 The same cluster of trajectories as in Figs. 1 and 2 is represented in a trajectory wall

till about 10:00. There was also a smaller congestion on the southeast from about 6:20 till about 7:20 in the morning. We also make a more general notice that in the morning the speeds were quite low on almost the whole length of the route. In the afternoon, a short period of obstructed traffic occurred from about 15:30 till about 17:00. This is also consistent with our previous observations (Fig. 1).

The mouse pointing gives us temporal information only on the elementary level of a single trajectory segment. To enable a higher level of temporal analysis, the display includes an element called time lens, which is visible in the lower right corner of Fig. 3. The time lens shows temporally aggregated information for an interactively defined spatial query area (a circle of a chosen radius around the mouse cursor position, visible in the lower left corner of Fig. 3). The interior of the time lens shows the relative spatial positions of the trajectory points within the selected area. The points are represented by dots coloured according to the attribute values. The ring of the time lens represents one of the temporal cycles: 4 quarters of a year, 12 months of a year, 7 days of a week, or 24 h of a day. In our example, the daily cycle is chosen. The ring is divided into bins corresponding to the units of the chosen cycle (hours in our case). The fill levels of the time bins visualize temporally aggregated information about the trajectories that intersect with the query area. The possible aggregates are the count of the trajectories, the total time spent in the query area (i.e., the sum of the times from all trajectories), and the average time (i.e., the total time divided by the count of the trajectories). In our example, the time lens shows the average times. The division of the bin contents into coloured segments shows the proportions of attribute values from different value intervals within the aggregates.

With the mouse, we have selected a query area in the place on the northwest where there are many trajectory segments with low speed values. The time lens shows us that such values mostly occurred in the hours from 5 till 9 in the morning (i.e., the intervals from 5–6 o'clock to 9–10 o'clock). Furthermore, we see great differences in the average times spent in the query area in different time intervals of the day. In the morning hours from 5 to 7, it took from 16 to 25 min on average for a car to move through the query area. In the next two hours, the average times decreased to 9 and 8 min, while during the rest of the day, the average times were 2–3 min. Hence, during the traffic congestion, the car drivers lost on average from 6 to 23 min of their time in comparison to normal driving.

3 Conclusion

In this extended abstract, we address an important and challenging problem of the visualization of positional attributes of trajectories in space and time. To overcome the problems caused by overlapping and intersecting trajectories, we propose to transform time into sequential order. This procedure transforms the space–time cube into a trajectory wall display.

We demonstrate by example that the proposed transformation enables effective and efficient visualization of positional attributes and facilitates interactive pattern detection. The technique can be applied to numeric attributes (with classification by given class intervals) and to qualitative variables. It works well with moderate-size data sets, up to several thousands of trajectories. To apply it to larger data, it is necessary to combine the trajectory wall displays with other interactive tools such as temporal animation, clustering, queries etc. (Andrienko and Andrienko 2013; Andrienko et al. 2013).

To enable detection of more complex patterns, it is essential to support flexible ordering of trajectories in the stack. Our current implementation allows ordering by different aspects of time (chronological order, hours of day, hours of week etc.) and by arbitrary attributes such as median or maximal speed. As a direction for improvement, we consider ordering trajectories by times of passing a selected geographical region.

References

Andrienko N, Andrienko G (2006) Exploratory analysis of spatial and temporal data. A systematic approach. Springer, Berlin
Andrienko N, Andrienko G (2013) Visual analytics of movement: an overview of methods, tools, and procedures. Inf Vis 12(1):3–24
Andrienko N, Andrienko G, Gatalsky P (2003). Visual data exploration using space–time cube. In: Proceedings of the 21st international cartographic conference, International Cartographic Association, Durban, 10–16 Aug 2003, pp 1981–1983

Andrienko G, Andrienko N, Wrobel S (2007) Visual analytics tools for analysis of movement data. ACM SIGKDD Explor 9(2):38–46

Andrienko G, Andrienko N (2010) Dynamic Time Transformation for Interpreting Clusters of Trajectories with Space-Time Cube. IEEE Visual Analytics Science and Technology (VAST 2010) Proceedings, IEEE Computer Society Press, pp 213–214

Andrienko G, Andrienko N (2011) Dynamic Time Transformations for Visualizing Multiple Trajectories in Interactive Space-Time Cube. International Cartographic Conference (ICC 2011) Proceedings

Andrienko G, Andrienko N, Bak P, Keim D, Wrobel S (2013) Visual analytics of movement. Springer, Heidelberg

Hägerstrand T (1970) What about people in regional science? Paper Reg Sci Assoc 24:7–21

Harrower M, Brewer CA (2003) Colorbrewer.org: an online tool for selecting colour schemes for maps. Cartogr J 40(1):27–37

Kapler T, Wright W (2005) GeoTime information visualization. Inf Vis 4(2):136–146

Kraak M-J (2003) The space–time cube revisited from a geovisualization perspective. In: Proceedings of the 21st international cartographic conference, International Cartographic Association, Durban, 10–16 Aug 2003, pp 1988–1995

Rinzivillo S, Pedreschi D, Nanni M, Giannotti F, Andrienko N, Andrienko G (2008) Visually-driven analysis of movement data by progressive clustering. Inf Vis 7(3/4):225–239

Schreck T, Bernard J, Tekusova T, Kohlhammer J (2008) Visual cluster analysis in trajectory data using editable Kohonen maps. In: Proceedings IEEE symposium on visual analytics science and technology (VAST 2008), IEEE, pp 3–10

Slocum TA, MacMaster RB, Kessler FC, Howard HH (2009) Thematic cartography and geovisualization, 3rd edn. Pearson Education, Upper Saddle River

Tominski C, Schumann H, Andrienko G, Andrienko N (2012) Stacking-based visualization of trajectory attribute data. IEEE Trans Vis Comput Graph (Proc IEEE Inf Vis 2012) 18(12):2565–2574

Visual Analysis of Lightning Data Using Space–Time-Cube

Stefan Peters, Hans-Dieter Betz and Liqiu Meng

Abstract This paper describes a framework for a visual analysis of lightning data described by 3D coordinates and the precise occurrence time. First lightning cells are detected and tracked. After that we developed a GUI (interactive graphic user interface) in order to enable the visual exploration of movement patterns and other characteristics of lightning cells. In particular we present different visual concepts for the dynamic lightning cells and tracks within a Space–Time-Cube and a 3D view. Furthermore a statistical analysis is presented. The developed GUI which aims to support decision making includes the visual and statistical representation of cell features as centroid, extension, density, size etc., within a specific temporal and spatial range of interest.

1 Introduction

In the last years lightnings became of interest of scientists not only due to their associations with thunderstorms but also because of their responsibility in atmospheric chemistry (Betz et al. 2009). To provide a reliable insight into the dynamics of lightning data for decision makers, lightning cells, tracks and their features have to be represented in the most suitable way. Mackaness et al. (2007) declared that the strength of maps is their capability to abstract geographic

S. Peters (✉) · L. Meng
Technical University Munich, Department of Cartography, Munich, Germany
e-mail: stefan.peters@bv.tum.de

H.-D. Betz
nowcast GmbH, 81377 Munich, Germany

information at various scales. Consequently different patterns can be exposed and features belonging to the respective spatial phenomena can be visualized. Today's information society has to deal with an increasingly exceeding amount of data, thus the demand for data abstraction is more and more needed.

Andrienko et al. (2008) stated that geovisualization aims to effectively present abstract information which expose patterns that cannot be recognized directly in the landscape. Investigations in this field rely on expertise in exploratory data analysis and visual analytics, but also in cognitive ergonomics and interface design. Exploratory tools for visual analysis of point data can be found within previous studies as in Krisp et al. (2009), Krisp and Peters (2010), Krisp et al. (2010) and Peters and Krisp (2010). Investigations in this paper may support visual analytics and data analysis approaches. MacEachren and Kraak (2001) already emphasizes that it is needed to tap the full potential of technological advances that permit not only complex multifaceted displays and dynamically linked views, but also take benefit of non-visual perceptual channels. Virrantaus et al. (2009) stated that Geovisualization techniques have extended the map medium to embrace dynamic, three- and four-dimensional data representation.

While exploring geodata which represent dynamic phenomena, efficient visual exploration might enable appropriate data analysis. However, when dealing with complex and large data sets, up to date visual analysis methods are often not very efficient. Common problems are in particular visual disorder, significant overplotting and thus data unreadability (Andrienko et al. 2008). To detect time dependent location based events, research on dynamic patterns is used among others. Questions to be solved are where and when the event takes place. In comparison to changes, events correspond to a higher knowledge level, and thus events are more important for decision makers. Besides, some pattern behaviors might be more complex and even irregular. However, the visualization of these patterns can be of utility for supplementary visual explorations (Krisp et al. 2012). In this study, event refers to the occurrence and dynamics of thunderstorms represented by lightning cells. Within this paper we investigate an explorative method for visual analyzing 3D lightning data with the aim to detect and understand spatial–temporal patterns of moving thunderstorms.

2 Methodological Framework: State of the Art

2.1 Lightning Data Detection and Position Accuracy

The three-dimensionally resolved total lightning data used in our experiment are provided by the European lightning detection network LINET (Betz et al. 2009). LINET has been developed at the Physics Department of the University of Munich and put into continuous operation by nowcast GmbH in 2006. In 2012 the network comprised 130 sensors in 30 European countries. The 3D-feature to report the

emission altitude of in-cloud (IC) strokes is unique. Due to optimized time-of-arrival techniques the 2D-location accuracy of cloud-to-ground (CG) strokes amounts to about 150 m. IC altitudes have an average accuracy of about 10 %, i.e. the emission point of an IC stroke in 10 km height above ground is determined to ±1 km. Higher accuracy is not needed for the data users; what matters is the time evolution of emission altitudes as a function of storm conditions. For example, when a thunderstorm grows to a severe weather cell, the IC-heights increase because the storm also grows vertically.

The location accuracy of LINET has been tested in many ways. Strikes into towers of known position verified the horizontal 2D-resolution of better than 150 m. It is more intricate to check the altitudes. A first test consisted of comparisons with 3D weather radar data: the IC locations coincided very well with the relevant convective cores. In more detailed comparison a LINET system has been set up at KSC (Kennedy Space Centre) in Florida, where NASA operates a special 3D-lightning detection network exploiting VHF (very high frequencies). It turned out that the IC information from LINET matched perfectly with the well resolved 3D-NASA data. Thus, the reported LINET IC heights can be used for all kind of evaluation purposes.

For tracking lightning cells, first cells have to be identified based on spatial–temporal clustering of the detected lightning point data. Usually the time intervals of 10–15 min are applied to divide lightning point data sets into temporal frames. If lightning cells are identified within each time interval, connected cells can be allocated to enable cell tracking. Thus appearance, changes within size/shape/location/density, merging, splitting and disappearing of each cell can be investigated.

2.2 State of the Art of Cell Identification and Tracking

To detect and track lightning cells is one of the tasks of thunderstorm investigations. Thunderstorm nowcasting algorithms consist of three steps: cell identification, cell tracking, and cell prediction. This work is based on existing methods for lightning cell identification and lightning cell tracking.

There are a number of methods for thunderstorm cell identification, tracking and nowcasting based on satellite data, radar data, lightning data, microwave-based temperatures or a combination of them. Meyer (2010) provides a good overview about existing thunderstorm tracking and nowcast methods. Li et al. (1995), Johnson et al. (1998) and Zinner et al. (2008) introduce methods based on satellite image data. Dixon and Wiener (1993), Handwerker (2002) and Hering et al. (2004) describe solutions based on radar data. The methods of Steinacker et al. (2000) and Bonelli and Marcacci (2008) can handle radar data and lightning data in combined way. Betz et al. (2008) introduced a cell tracking approach based only on lightning data derived from the established LINET network.

Cell tracking techniques can be distinguished into pattern-oriented correlation techniques and overlapping techniques. In this work we use an overlapping technique

proposed in Zinner et al. (2008) and Hering et al. (2004). Cell tracking methods based on cost functions were used in Bonelli and Marcacci (2008) and Dixon and Wiener (1993). The objective thereby is to obtain the most plausible cell assignments. A Gaussian filter is used in several cell tracking approaches to remove small cells. Furthermore certain allocation restrictions are defined based on estimated wind data or velocity vectors. Some of the tracking approaches consider the cell splitting and merging behavior.

Herd or crowd movement theories are of limited use for lightning cell analysis. Lightnings can be explained as moving collectives within a dynamic thunderstorm. Nevertheless, they cannot be interpreted as discrete dual-aspect phenomena like crowd events are Galton (2005).

2.3 Lightning Cell Visualization

In all previous works, lightning point data, detected lightning cells and derived cell tracks are visualized in 2D. Lightning cells are either displayed as 2D convex hulls with or without the underlying lightning point data. Due to recent improvements of lightning data detection and accuracy, there is a growing demand on multidimensional and interactive visualization in particular for decision makers. The idea for an interactive explorative tool for lightning data had been already introduced by Peters et al. (2013). This work will introduce an extended framework for an interactive graphic interface with diverse methods for visually exploring the lightning cells and tracks.

3 Test Dataset of Lightning Points

Our test dataset consists of all lightning points detected by LINET (and provided by nowcast GmbH) during July 22, 2010, in the region of Upper Bavaria, Germany, with the geographic extension of:

Latitude = 47–50° North, Longitude = 10–13° East.

From 1 pm until midnight a thunderstorm crossed Upper Bavaria where altogether 34809 lightning points were captured. Every lightning point is an individual discharge (also called stroke) and LINET provides the 3D location (Lat, Lon, Altitude) and the exact stroke time in Table 1.

Table 1 Example of a stroke

Date	Time	Latitude (°)	Longitude (°)	Altitude (km)
22.07.2010	14:05:22.7138192	51.0142	4.4367	11.2
22.07.2010	14:05:22.4278749	51.0018	4.4497	0

Figure 1 shows the entire lightning data set (blue dots) in 3D and 2D. The red lines on the ground illustrate the Upper Bavaria region and its districts. Furthermore the red dot represents the location of Munich Airport (MUC). The vertical red line underlines the MUC location in 3D.

For lightning analysis and nowcasting the visual presentation given in Fig. 1 is very limited. Lightning cells can not being identified, much less distinguished, time information is missing and thunderstorm movement direction can't be recognized either.

With our approach we try to provide visual different analytic tools for data experts and decision makers to overcome these shortcuts and to get as much visual information as possible out of a lightning dataset.

4 Development of an Interactive Tool for Lightning Cell Analysis

The workflow of lightning cell analysis is illustrated in Fig. 2. The test data described in previous section are first divided into cloud-ground lightning (CG) and inter-cloud-lightning data (IC). To remove outliers, IC data were confined to points with an altitude between 2 and 18 km.

In the second step, lightning point data are segregated into 10 min time intervals.

In the next step all points within each time interval are grouped into cells. Point clustering can be classified into 3 major types: Density-based, Hierarchical and Partitioning methods. An overview about existing point clustering methods is provided by Jain and Dubes (1988). We used a simple partitioning method. Most partitioning methods cluster objects based on the distance between objects. Such
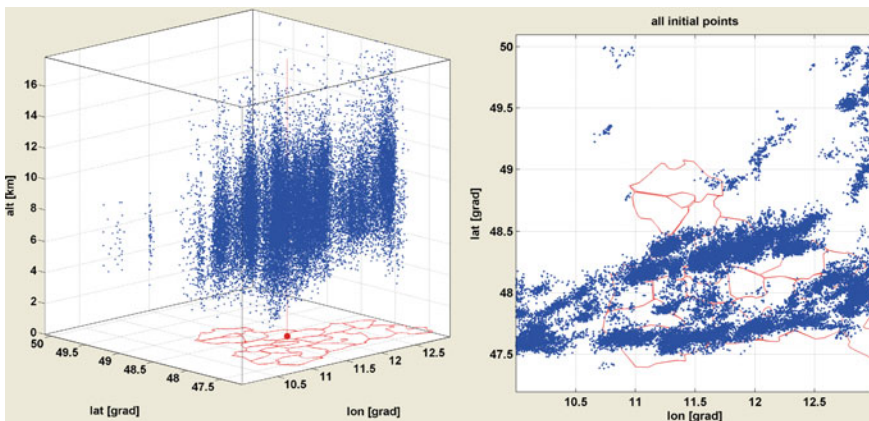
Fig. 1 Lightning data of 22.07.2010 over Upper Bavaria (*left* 3D, *right* 2D)
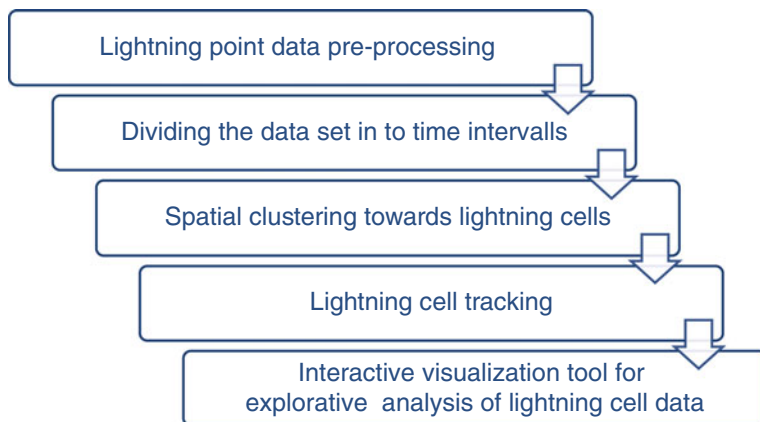
Fig. 2 Workflow scheme of lightning data analysis

methods can find arbitrary shaped clusters. The general idea here is to expand the given cluster as long as the distance towards a neighborhood point does not exceed some threshold. Such methods can be used to filter out noise or outliers (Han and Kamber 2006). We clustered all lightning data belonging to the same time interval based on overlapping point buffers of 6 km radius. The resulting clusters represent lightning cells. A minimum cell number size of 10 points was set. This threshold could be increased to remove small lightning cells.

For cell tracking, cells which overlap within 2 moments of time were detected and allocated. The approach is based on the tracking method of Zinner et al. (2008) and Hering et al. (2004). Thus cell splitting and merging are considered as well.

After having tracked the cells an interactive graphic user interface (GUI) was designed to provide an explorative analysis of the dynamic lightning cell data. Figure 3 illustrates 6 different presentations of a 3D lightning cell: (a) lightning cell points, (b) cell outlines/cuboid and cell centre, (c) density-based cell centre with a spherical radius indicating the scope, (d) ellipsoid fitted into the 3D cell extension, (e) 3D convex hull, (f) combination of b, c and e.
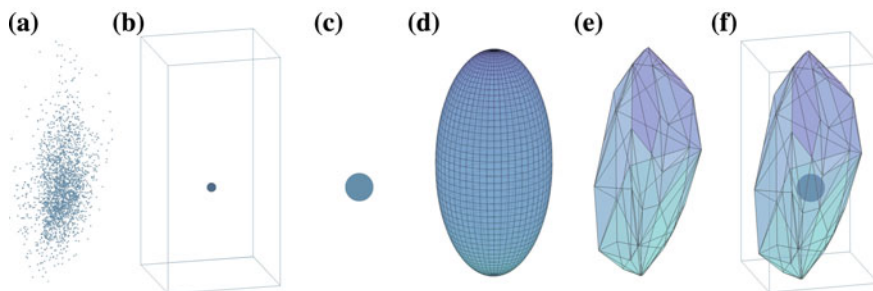
Fig. 3 Different options for lightning cell visualization
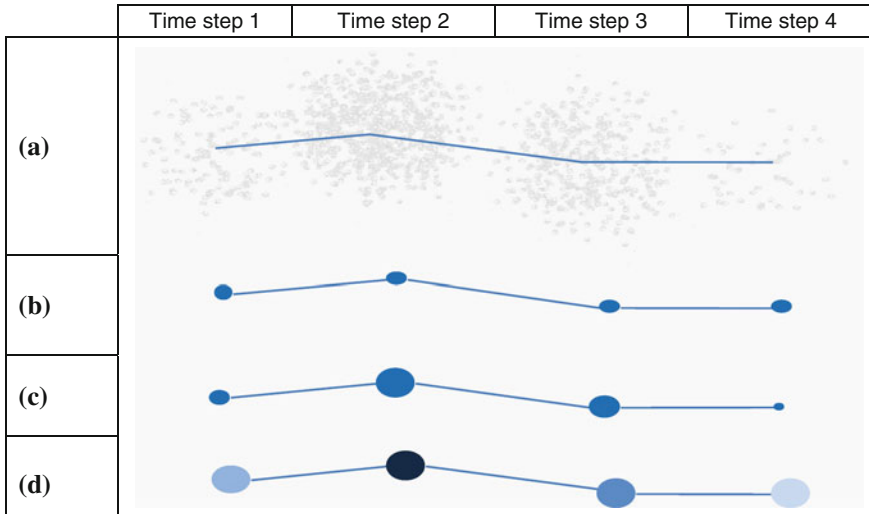
Fig. 4 Different options for cell-tracking visualization

Figure 4 shows 4 different 3D cell track presentations for an example with 4 time steps: (a) track line, connecting the point density based cell centres, lightning point data are displayed in grey in the background, (b) track line with point density based cell centres, (c) track line with point density based cell centre sizes, (d) track line with point density based cell centre colors.

4.1 Space–Time-Cube for Lightning Data

Figure 5 left illustrates allocated lightning cells (over Upper Bavaria) over 1 entire day (22.07.2010) whereas on the right hand site the detected cells of only 1 time frame is shown. All cells allocated to the same track are distinguished by semi-transparent colors. Two larger tracks can be identified (purple and green track cells/lines). From these static maps the following information cannot be extracted: which direction did a lightning cell moved? To which specific time interval refers a certain cell?

Another disadvantage is that while plotting cells of several time intervals (as in Fig. 5 left) cells occur overlapped and changes of cell sizes, cell shapes of cell densities (number of lightning points) are very hard to identify visually. To overcome these shortcomings an interactive Space–Time-Cube was investigated. An overview about concept, applications and improvements while using Space–Time-Cube for geodata visual analysis is provided by Kraak (2003).

Figure 6 shows the investigated concept of how lightning cells can be visualized within a Space–Time-Cube: (a–c) cells of one specific time interval: (a) cells
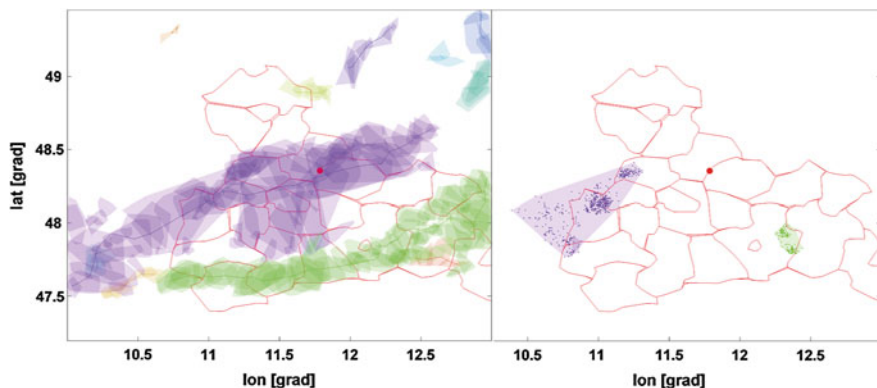
Fig. 5 *Left* lightning cells over 1 day, *right* lightning cells over 10 min

distinguished by track-color, (b) cells including lightning points, (c) cells with point density based cell centre size, (d–f) cells of time interval series: (d) cell tracks distinguished by track-color, (e) cells with track line and lightning points, (f) cells with track line and point density based cell centre size. The color brightness of the cell centre sphere can also reflect the density differences between cells.

Trough an interactive GUI, these Space–Time-Cube data can be explored from different perspectives (from different angles/distances) using zoom, pan, rotation tools. Besides a time slider allows to animate and to discover the cell dynamics within a certain time frame. A GUI also enables to switch between or to combine different visualization concepts for lightning cells and cell-tracks demonstrated in Figs. 3, 4, 6. Furthermore an examination from above (2D-view) enables the same view and thus it provides the same information like the traditional plot as shown in Fig. 5.
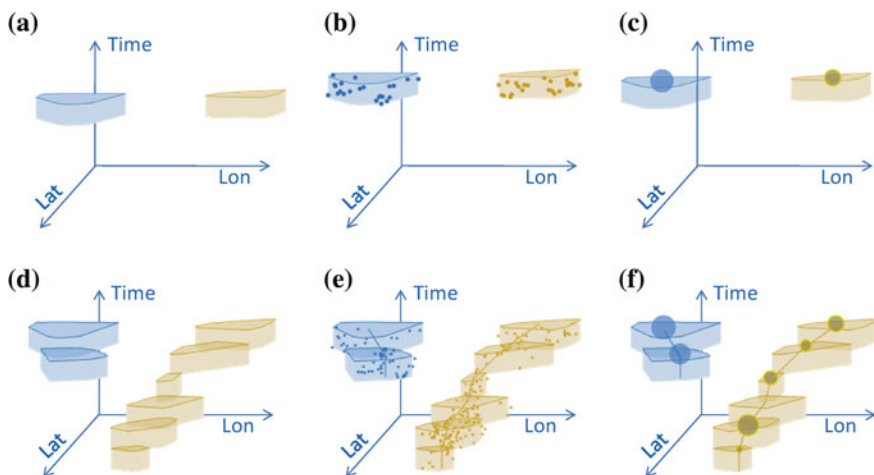
Fig. 6 Concept of space–time-cube

Table 2 Overview about statistical analysis of lightning cell/track data

Data for each lightning cell	Data for each lightning track	Correlation between
Density based centre	Lifetime (nr. of time intervals)	Cell density and altitude
Cell extension (2D/3D)	Variation of cell forms	Cell volume and density
Cell density (number of lightning points)	Variation of cell densities	Cell lifetime and altitude
Area (2D-projection)	Variation of cell areas	Cell form and density
Volume (of 3D convex hull)	Variation of cell volumes	Cell velocity and lifetime
min, max, average altitude	Variation of cell altitudes	
Velocity	Variation of cell velocities	

4.2 Statistical Analysis of Lightning Data

Also essential for lightning cell analysis are statistical data derived from the processed cell tracking. Table 2 gives an overview about important statistical data which could be of interest. These statistical data could be included in the lightning GUI, thus the respective data of the current view are illustrated in an appropriate way (e.g. by using diagrams, highlighting the relevant cell/track).

5 Results and Evaluation

The designed visualization concepts for dynamic lightning cells were implemented in an interactive GUI for the given test dataset. Figure 7 illustrates the results on a screenshot of that GUI: two synchronized plots are created, left: Space–Time-Cube (z-axis: time in minutes), right: 3D plot of the lightning data. Lightning data points can be loaded as text files containing only the point positions. Within the load-function allocated lightning cells will be calculated. On the left hand site explorative visual methods can be applied and combined by switching on the check-boxes: lightning cells can be displayed by the point convex hull, as ellipsoid, by the cuboid outlines or as spheres—located at the cell centre. The cell centre can be visualized with a point density based sphere size and also with a point density based sphere color. The provided cell and track graphics data can be shown in a 3D plot, in a 2D plot and also in combination at the same time: 3D plus projected 3D on the 2D surface.

Currently only the clustered lightning points of the entire test data set are shown in Fig. 7 whereby cell-tracks are distinguished by different colours. The advantage of the Space–Time-Cube is providing the missing information in the 3D plot on the right hand site (which can be changed by rotation towards a 2D plot). The 3D plot gives altitude information, but time and movement direction information are missing. An advantage comparing to the 3D plot (Fig. 7 right) the Space–Time-Cube plot avoids overlapping of cells and thus improves clarity and facilities data interpretation.
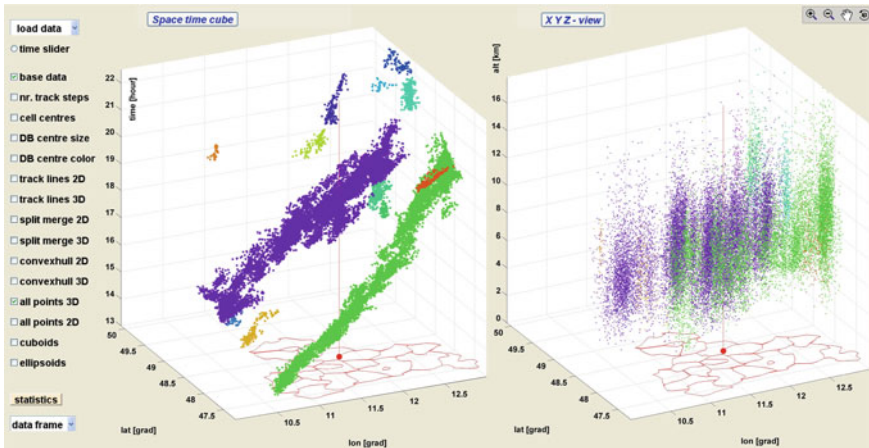
Fig. 7 Interactive GUI for visual analysis of lightning test data set, 22.07.2010

The combination of both plots is complementary and provides through its interactive user interface a variety of information about the lightning cell dynamics. In the Space–Time-Cube two main lightning cells (green and purple) can be clearly identified and also when and with which density a cell is at which location. It can be seen also in Fig. 8 that the thunderstorm represented by the purple lightning cell crosses Munich airport (red dot and vertical line) at about 7 pm. Figure 8a additionally displays cell track lines, number of cells per track and lightning point density based cell centre sizes. Merging and splitting can be recognized from the
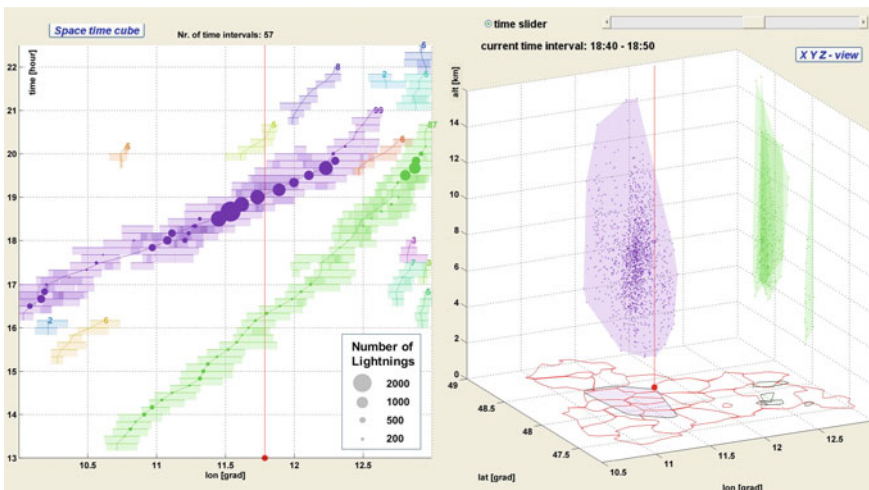
Fig. 8 a Space–time-cube plot. **b** 3D plot of 1 time step (18:40–18:50)

Space–Time-plot as well. The GUI also enables an interactive presentation of statistical data as described in "Austro–Hungarian North Pole Expeditionand its Role Model". Thus the user can interactively request statistical values, for example cell densities, cell area or cell volumes.

Furthermore the data frame can be limited to concentrate on specific areas of interest. Besides rotation, pan, and zoom tools enable the interactive discovering. By enabling the time slider, specific time intervals can be investigated by plotting and animating. Figure 8b shows a selected time interval (18:40–18:50) containing the densest cell in 3D and a 2D projection: the purple convex hull includes about 2000 lightning points.

6 Conclusion and Outlook

This work presented a practical framework for visual analysis of lightning data. Lightning cells were at first extracted from lightning point data. An interactive graphic user interface was then developed to enable the investigation of the dynamics of the lightning cells. Thereby a comprehensive set of visual explorative methods, e.g. 3D view, Space–Time-Cube, time slider tool, presentation of results of statistical data analysis, was introduced and implemented. In future work, a nowcasting will be conducted and visualized. The idea is to predict the following cell features for the next 10–60 min including location, centre, extension, density, area, volume, lifetime and feature probabilities. The main focus will be set to a suitable interactive visualization of the predicted featured within the GUI. Furthermore, user tests, in particular tests with lightning experts and interested lay, will be implemented to verify and improve the GUI usability.

References

Andrienko G, Andrienko N, Dykes J, Fabrikant SI, Wachowicz M (2008) Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. Inf Vis 7(3–4):173–180

Betz HD, Schmidt K, Oettinger WP, Montag B (2008) Cell-tracking with lightning data from LINET. Adv Geosci 17:55–61

Betz HD, Schmidt K, Oettinger WP (2009) LINET—an international VLF/LF lightning detection network in Europe. In: Betz HD, Schumann U, Laroche P (eds) Lightning: principles, instruments and applications. Springer, Dordrecht, pp 115–140

Bonelli P, Marcacci P (2008) Thunderstorm nowcasting by means of lightning and radar data: algorithms and applications in northern Italy. Nat Hazards Earth Syst Sci 8(5):1187–1198

Dixon M, Wiener G (1993) TITAN: thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology. J Atmos Ocean Technol 10(6):785–797

Galton A (2005) Dynamic collectives and their collective dynamics. In: Cohn AG, Mark DM (eds) Spatial information theory. Springer, Berlin, pp 300–315

Han J, Kamber M (2006) Data mining: concepts and techniques. Morgan Kaufmann, San Francisco

Handwerker J (2002) Cell tracking with TRACE3D—a new algorithm. Atmos Res 61(1):15–34

Hering A, Morel C, Galli G, Sénési S, Ambrosetti P, Boscacci M (2004) Nowcasting thunderstorms in the Alpine region using a radar based adaptive thresholding scheme. In: Proceedings of ERAD

Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice-Hall, Inc., Upper Saddle River

Johnson J, MacKeen PL, Witt A, Mitchell EDW, Stumpf GJ, Eilts MD, Thomas KW (1998) The storm cell identification and tracking algorithm: an enhanced WSR-88D algorithm. Weather Forecast 13(2):263–276

Kraak MJ (2003) The space–time cube revisited from a geovisualization perspective. In: Proceedings of 21st international cartographic conference, pp 1988–1996

Krisp J, Peters S (2010) Visualizing dynamic 3D densities: a Lava-lamp approach. In: 13th AGILE international conference on geographic information science

Krisp JM, Peters S, Murphy CE, Fan H (2009) Visual bandwidth selection for kernel density maps. Photogrammetrie Fernerkundung Geoinf 5:445–454

Krisp J, Peters S, Burkert F, Butenuth M (2010) Visual identification of scattered crowd movement patterns using a directed kernel density estimation. In: SPM2010 mobile Tartu

Krisp JM, Peters S, Polous K, Fan H, Meng L (2012) Getting in and out of a taxi: spatio-temporal hotspot analysis for floating taxi data in Shanghai. In: Networks for mobility 2012, Stuttgart

Li L, Schmid W, Joss J (1995) Nowcasting of motion and growth of precipitation with radar over a complex orography. J Appl Meteorol 34(6):1286–1300

MacEachren AM, Kraak MJ (2001) Research challenges in geovisualization. Cartogr Geogr Inf Sci 28(1):3–12

Mackaness WA, Ruas A, Sarjakoski LT (2007) Generalisation of geographic information: cartographic modelling and applications. Elsevier Science, Amsterdam, p 386

Meyer V (2010) Thunderstorm tracking and monitoring on the basis of three-dimensional lightning data and conventional and polarimetric radar data. In: DLR, Deutsches Zentrum für Luft- und Raumfahrt. p 128

Peters S, Krisp JM (2010) Density calculation for moving points. In: 13th AGILE international conference on geographic information science

Peters S, Meng L, Betz HD (2013) Analytics approach for lightning data analysis and cell nowcasting. In: EGU general assembly conference abstracts, pp 32–33

Steinacker R, Dorninger M, Wölfelmaier F, Krennert T (2000) Automatic tracking of convective cells and cell complexes from lightning and radar data. Meteorol Atmos Phys 72(2):101–110

Virrantaus K, Fairbairn D, Kraak M-J (2009) ICA research agenda on cartography and GI science. Cartogr J 46(2):63–75

Zinner T, Mannstein H, Tafferner A (2008) Cb-TRAM: tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data. Meteorol Atmos Phys 101(3):191–210

Silhouette-Based Label Placement
in Interactive 3D Maps

Christine Lehmann and Jürgen Döllner

Abstract This paper presents a silhouette-based technique for automated, dynamic label placement for objects of 2D and 3D maps. The technique uses visibility detection and analysis to localise unobstructed areas and sihouettes of labeled objects in the viewplane. For each labeled object, visible silhouette points are computed and approximated as a 2D polygon; the associated label is finally rotated and placed along an edge of the polygon in a way that sufficient text legibility is maintained. The technique reduces occlusions of geospatial information and map elements caused by labels, while labels are placed close to labeled objects to avoid time-consuming matching between legend and map view. It ensures full text legibility and unambiguity of label assignments by using actually visible 2D silhouette of objects for label placement. We demonstrate the applicability of our approach by examples of 3D map label placement.

1 Introduction and Related Work

Label placement or labeling is an old discipline that has its origin in 2D cartography (Imhof 1985). Different labeling strategies have been devised to attach textual information to objects. A large number of labeling strategies exist for point feature labeling of 2D objects (Bekos et al. 2010; Christensen et al. 1994; Mote 2007; Wagner et al. 2001) and 3D objects (Stein and Décoret 2008; Maaß and Döllner 2006b). The challenge in point feature labeling comes from the high density of labeled objects or Point-of-Interests (Kern and Brewer 2008).

C. Lehmann (✉) · J. Döllner
Computer Graphics Systems, Hasso-Plattner-Institute, University of Potsdam, Potsdam, Germany
e-mail: christine.lehmann@hpi.uni-potsdam.de

Algorithms commonly use external lines to connect label and labeled object, which is called *external labeling*. Fewer strategies exist for line feature labeling (Maaß and Döllner 2007) and area feature labeling (Maaß and Döllner 2006a), which overlay objects with labels. Such approaches are referred to as *internal labeling* approaches.

From cartography, general principles and requirements on text labels are well-known (Imhof 1985), including *legibility*, *clear graphic association*, and, in particular, *minimum covering* of map content. In particular, the last issue represents a major factor that affects the clarity of the visualization (Christensen et al. 1994). With internal labeling, much content is covered by labels. With external labeling, only one point of the labeled object is required. In case of illustrative labeling examples, i.e., labels can be placed exclusively around the content and thus provide less covering of labeled objects. However, for general labeling examples, external lines can also cover adjacent labeled objects.

In contrast to *static* label placement, *dynamic* label placement requires that labeling algorithms are suitable for responsive visualization systems and interactive applications (Kopetz 1993) and thus perform all computations fast enough. Furthermore, interactive applications are characterized by frequent changes in view and—in case of 3D visualizations-also by changes in perspective. This makes it more difficult for the user to focus and perceive the desired information, compared to a static map, for example, where the user has more time to study the given single view. In the field of human user interfaces, this refers to the task of the user to keep the "locus of attention" (Raskin 2000) in interactive applications, i.e., to detect and interpret the difference between two consecutive views of the visualization. Hence, if there is too much information visualized, the user looses the locus of attention and consequently does not perceive the message of the visualization anymore. As a conclusion, our labeling method focuses on providing an appropriate amount of information to the user using visibility-driven label placement (Lehmann and Döllner 2012). Using visibility information of labeled objects, which we call *visibility-driven* label placement, is essential for dynamic label placement, as it reduces missing labels and label cluttering (Lehmann and Döllner 2012).

In this paper, we present a novel labeling strategy, called *silhouette-based label placement* that provides high visibility of underlying map content without losing spatial closeness between label and labeled object. Labels are placed along the visible silhouette of labeled objects (Fig. 1). In this sense, silhouette-based labeling is a balance between external and internal labeling. Our method is fully automated and does not require manual refinement. The label placement is renewed for each frame and reacts on interactive frame rates. Hence, it is suitable for interactive visualization applications.

We will further present how text and character transformations, which are necessary to align labels to visible silhouettes, can be efficiently implemented. With our font rendering method, we overcome legibility problems caused by aliasing artifacts and enable dynamic transformations of text and single characters.

Fig. 1 Occlusions of labeled objects caused by labels are reduced to preserve their information, while fast label assignment is enabled

Our GPU-based rendering method for text labels can be easily integrated into existing rendering applications, e.g., into server-side applications that need text visualizations, as the main part of text rendering is placed in a pixel shader. We demonstrate the applicability of our labeling method for 2D and 3D visualization applications by city and indoor labeling examples.

2 Silhouette-Based Label Placement

For each labeled object, pixels of the visible area are detected. As this detection is performed in the image space, where both 2D and 3D objects are represented by 2D regions, our method can be applied for 2D and 3D maps. Visible regions of 3D objects can consist of separate parts, for instance caused by occluding foreground objects. These parts or clusters in the visible point set are detected using flat Euclidean clustering (Jain et al. 1999). The cluster containing the highest number of points is selected for further processing (Fig. 2). From the cluster, we extract the boundary points or *the visible silhouette*. For each boundary, the vector pointing to the next inner point is computed, which we call the *inner vector* in this context. Hence, we call the resulting set of inner vectors for a given point set the *inner vector field*. In the following, we present our completely automated technique for aligning labels to visible edges of arbitrary 2D and 3D objects.
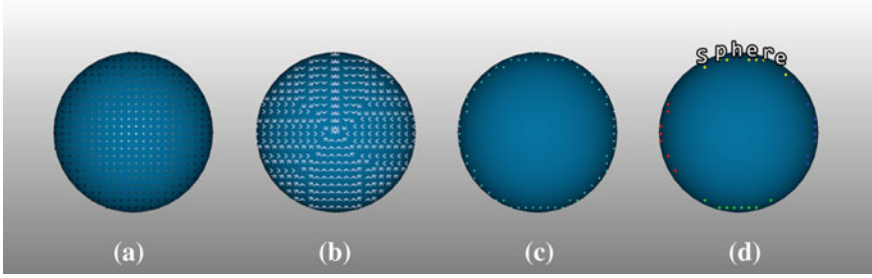
Fig. 2 Visible regions of labeled objects (example: spherical object) are analyzed in image space. **a** The *distance field* describes the distance for each visible point to the boundary. Points with highest distance are depicted white. Using the distance field, the visible silhouette (**c**) is selected. **b** The *inner vector field* describes the direction from each point to the inner points. It is used to detect *edges* in the visible silhouette (**d**), i.e., *left* edge (*red*), *bottom* edge (*green*), *right* edge (*blue*), and *top* edge (*yellow*). In this example, the label "Sphere" is aligned to the *top* edge

2.1 Detecting Visible Edges

We define an *edge* as a sorted list of 2D points that describes one side of the visible boundary of a 3D object. Thereby, the edge points are sorted in ascending order by their x-value. From the visible silhouette, i.e. the visible boundary points, we classify into top, bottom, left, and right edge (*edge classes*). Visible regions can have various shapes, for example, a shape closely matching a circle or a rectangle. The challenge is to find edges for arbitrary shapes of visible regions. This is difficult in case of shapes closely matching a circle, as a circle naturally has no edges or sides. However, we can analyze a boundary point relatively to the inner points of the shape. Hence, we use the inner vector field of the shape for edge detection.

To assign a point P_i to the correct edge class, the inner vector field is used: the point's inner vector N_i points to the inside of the visible region. Hence, we distinguish edges using the dot products $\langle N_i, E_0 \rangle$ and $\langle N_i, E_1 \rangle$, with $E_1 = (-1, 1)^T$ and $E_0 = (1, 1)^T$ (Fig. 3). For edge classification, only the sign of the dot product is relevant, thus, normalization of E_0 and E_1 is not required. Visible edges are classified using the following equations:

Bottom edge: $\langle N_i, E_0 \rangle > 0$ and $\langle N_i, E_1 \rangle > 0$.
Top edge: $\langle N_i, E_0 \rangle < 0$ and $\langle N_i, E_1 \rangle < 0$.
Left edge: $\langle N_i, E_0 \rangle > 0$ and $\langle N_i, E_1 \rangle < 0$.
Right edge: $\langle N_i, E_0 \rangle < 0$ and $\langle N_i, E_1 \rangle > 0$.

There are several ways to understand *bottom*, *top*, *left*, and *right* depending on the reference coordinate system. We can use the coordinate system defined by the eigenvectors of the visible point set, i.e., the 2D object or local coordinate system, or the axis-aligned coordinate system, i.e., the image space coordinate system. We

Fig. 3 To detect edges in the visible silhouette of labeled objects, the diagonal vectors e_1 and e_0 are used, i.e., the coordinate axes x and y rotated by 45°. Using the normal field and the diagonal vectors, points are assigned to one of the four regions *left*, *right*, *top*, and *bottom*
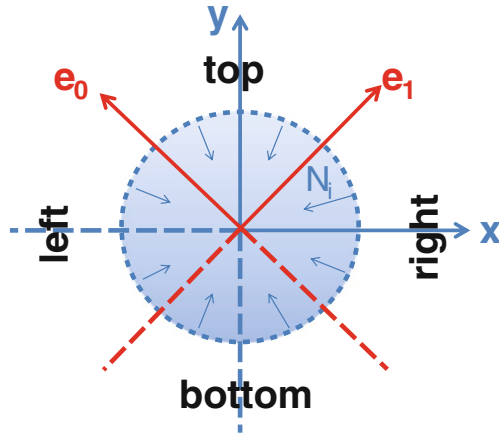
use the axes of the image space coordinate system to compute the absolute top, bottom, left, and right edge. When using a local coordinate system, the determination of common edges between adjacent labels is difficult, as edges cannot be compared without complex coordinate system transformation.

In case of concave shapes or shapes with holes, multiple edges can be detected for one edge class. In this case, the edges must be separated using flat clustering by the Euclidean Distance, i.e., the pixel distance between the edge points.

2.2 Selecting Edges

To select an edge, to which a given text can be aligned, we apply the following criteria:

1. The edge length in pixels must be sufficient for the given text length in pixels.
2. Horizontal edges (top and bottom edge) are preferred over vertical edges (left and right edge).
3. The last selected edge is preferred, if it has a sufficient length, so that spatial coherence between two consecutive views is improved.

Regarding criterion 1, the font size can be decreased to fit the text into the longest edge, if no edge is sufficiently sized for the text in original font size. According to cartography principles, objects of the same or similar type should obtain text labels with the same or similar font. Consequently, we assume that font size is constant for all labels, and blend out labels that do not fit to one of the four edges.

Regarding criterion 2, we set the following order for edge priorities in our examples: the highest priority is assigned to the top edge, followed by bottom edge, left edge, and right edge. The choice of edge priorities is related to reading

direction and, thus, depends on cultural aspects in text processing. For this reason, priorities are user-defined in our implementation.

Labels are placed along the boundary, but still inside the visible area of the labeled object. Hence, occlusions with other labels cannot occur, as visible areas are disjoint by nature. However, labels of adjacent visible areas can be placed closely, if the common edge is selected.

2.3 Computing Label Alignments

In the previous steps, we have computed the four visible edges, and selected one edge for each labeled object. In this section, we explain how the associated label is aligned to the selected edge, i.e., to the line sequence consisting of a tupel of 2D points or pixels. Single characters of a text label are translated separately by an offset vector. For the whole text, a list of offset vectors is computed. The start point of the text label is set to the start point of the edge, i.e., the furthest point on the left.

To generate the offset vector for the horizontally oriented edges (bottom and top edge), we collect all *relevant* edge points for each character (Fig. 4). An edge point is *relevant* for ith character, if it is located inside the character pixel bounding box $AABB_i$. From the relevant points, the average offset in y-direction is computed. For the vertically oriented edges (left and right edge), offsets are accordingly generated in x-direction.
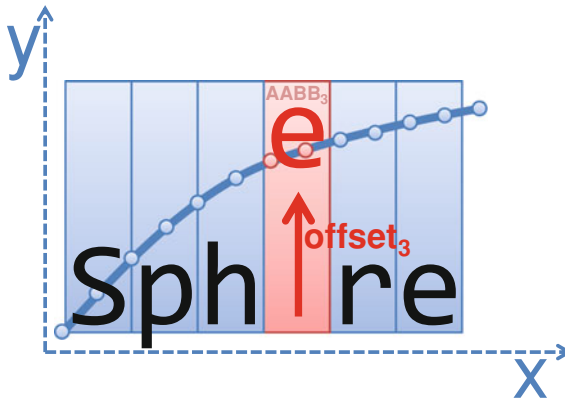
Fig. 4 Example offset computation: to align the character "e" of the label "Sphere" (3rd character) to the edge points (*blue* marked), the translation in y-direction, the *offset*, is computed. The **offset** is the average of the y-values of the edge points (*red* marked) inside the character bounding box *AABB*. The character position (x, y) of "e" (in image space) is then translated by the offset vector $(0, \textbf{offset})^{\textbf{\textit{T}}}$

Edge points do not necessarily need to be equidistant, if the sampling distance is higher than 1 pixel. Hence, a character can be located between two adjacent edge points without intersections. As a result, the translation offset between two adjacent characters can be very high, i.e., the character spacing between adjacent characters can be too loose. To enable sufficient character spacing for character translation in this case, we add the next adjacent edge points around the character bounding box. Additionally, we interpolate offsets using the arithmetic mean. As a result, we obtain a label alignment to a smoothed curve approximation of the edge.

The precision of the offset vector depends on the sampling distance. If the sampling distance is chosen too large, label alignments do not match anymore to the visible silhouette. Applying smaller sampling distances can lead insufficient performance. In our application scenarios with Full-HD resolution, we tested with a sampling distance of at least 10 pixels and maximally 50 pixels to provide reasonable label alignments with interactive frame rates.[1]

If the whole text was transformed, legibility could be reduced due to text distortion. Hence, to avoid artifacts in text rendering, we use a font rendering technique that enables character transformation, i.e., single characters can be transformed independently. However, we also allow transformations of the whole text by translation or rotation. To enable real-time text alignment to curves, we implemented a font rendering technique that completely runs on the GPU using a single pixel shader with the following input textures: two font textures as well as a text texture and an offset texture per label. As these textures are low-resolution textures, the total GPU memory consumption is also low. The pixel shader for character transforming and rendering is implemented using the OpenGL Shading Language (GLSL) (Rost 2006).

The offset vectors are passed to the pixel shader as an additional 1D texture (*offset texture*). In the pixel shader, each character is translated by the according offset contained within the offset texture.

3 Conclusions

We presented a visibility-driven, silhouette-based labeling technique for 3D virtual worlds such as 2D and 3D maps that reduces occlusions of labeled objects by labels while fast assignment by placing labels close to labeled objects is preserved. Our technique enables dynamic text transformations around the visible silhouette of labeled objects using GPU-based font rendering that preserves text legibility. Our method provides detailed user-defined configuration, including (1) the *edges priorities*, i.e., the bottom edge is always preferred against the other edges, (2) the *label distance to the visible silhouette*, i.e., labels can be aligned on the silhouette

---

[1] We tested on the following hardware: Intel Xeon CPU 2×2.6 GHz, NVIDIA GeForce GTX 480.
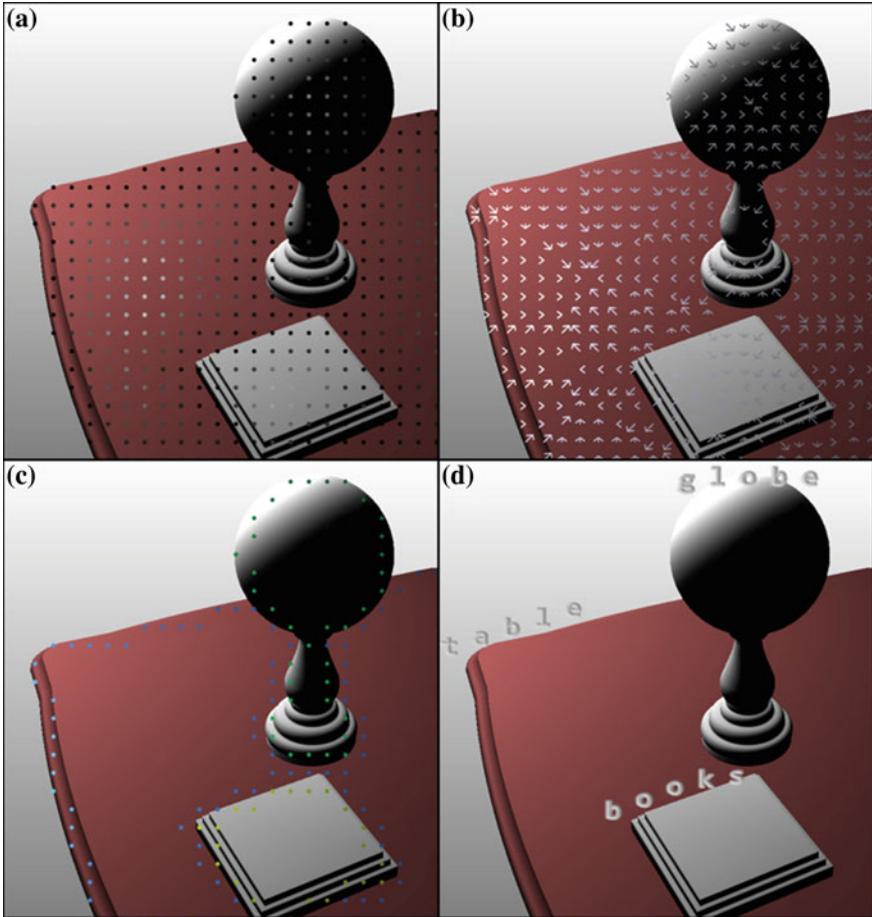
Fig. 5 Example for silhouette-based labeling applied to an indoor scene. From **a** to **d**: distance field, inner vector field, visible silhouettes, and aligned labels. In this example, we aligned labels to the *outer* visible silhouette of labeled objects and applied further smoothing to character translation

or to the inner/outer silhouette (Fig. 5), and (3) *label resizing*, i.e., labels are allowed/not allowed to be resized, if no edge is sufficiently sized.

Our labeling method projects labeled 3D objects into image space, i.e., 3D objects are represented as 2D areas in image space. Consequently, labeling of volumetric objects is reduced to labeling of areal objects (area feature labeling) to provide a generic approach for both 2D and 3D labeling. Kern and Brewer (2008) stated that area feature labeling is most difficult, so labels for areas must be placed first, followed by point and line feature labels. This ordering is mainly related to occlusions caused by point or line feature labels contained in area features, which are also labeled. With our approach, labels of objects, whose visible areas

degenerate to a point or a few points, are not depicted. They can be explored interactively by zooming in.

In our examples, we used a monospaced font type, i.e., all character glyphs have the same width. For a monospaced font type, the determination of offsets for edge alignment can be handled significantly more efficiently, in computational time and storage amount, than for a variable-width font type: for a variable-width font type, an additional offset texture is required that contains the character width information of the font type. Further, the pixel shader has to perform kerning and, thus, more texture look-ups. On the CPU, more complex distance computations for the offset vector have to be processed.

For future work, we want to compute the 2D skeleton of visible regions or *minor skeleton* (Freeman and Ahn 1984) respectively. Using the minor skeleton, we can align internal labels more precisely in the according visible regions. A real-time capable algorithm for image-based computation of the minor skeleton is provided by Telea and van Wijk (2002).

References

Abdi H, Williams L (2010) Principal component analysis. Comput Stat 2(4):433–459

Beauchemin SS, Barron JL (1995) The computation of optical flow. ACM Comput Surv 27(3):433–466

Bekos MA, Kaufmann M, Nöllenburg M, Symvonis A (2010) Boundary labeling with octilinear leaders. Algorithmica 57(3):436–461 (Scandinavian Workshop on Algorithm Theory)

Christensen J, Marks J, Shieber S (1994) An empirical study of algorithms for point feature label placement. ACM Trans Graph 14(3):203–232

Freeman H, Ahn J (1984) AUTONAP—an expert system for automatic map name placement. In: Internaional symposium on spatial data handling, pp 544–569

Imhof E (1985) Positioning names on maps. Cartogr Geogr Inf Sci 2:128–144

Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv 31(3):264–323

Kern J, Brewer C (2008) Automation and the map label placement problem: a comparison of two GIS implementations of label placement. Cartogr Perspect 60:22–45

Kopetz H (1993) Should responsive systems be event-triggered or time-triggered? Inst Electron Inf Commun Eng E76-D:1325–1332

Lehmann C, Döllner J (2012) Automated image-based label placement in interactive 2D/2.5D/3D maps. In: Symposium on service-oriented mapping

Maaß S, Döllner J (2006) Dynamic annotation of interactive environments using object-integrated billboards. In: 14th international conference on computer graphics, visualization and computer vision, pp 327–334

Maaß S, Döllner J (2006) Efficient view management for dynamic annotation placement in virtual landscapes. In 6th international symposium on smart graphics 2006, vol 4073. Springer, Heidelberg, pp 1–12

Maaß S, Döllner J (2007) Embedded labels for line features in interactive 3D virtual environments. In: AFRIGRAPH '07: proceedings of the 5th international conference on computer graphics, virtual reality, visualisation and interaction in Africa, pp 53–59

Mote, K. (2007). Fast point-feature label placement for dynamic visualizations. Inf Vis 6:249–260 (Data Structures and Algorithms)

Raskin J (2000) The humane interface: new directions for designing interactive systems. ACM Press, New York

Rost RJ (2006) OpenGL(R) shading language, 2nd edn. Addison-Wesley Professional, Boston

Stein T, Décoret X (2008) Dynamic label placement for improved interactive exploration. In: NPAR (symposium on non-photorealistic animation and rendering), pp 15–21

Telea A, van Wijk JJ (2002) An augmented fast marching method for computing skeletons and centerlines. In: Symposium on data visualisation

Vaaraniemi M, Treib M, Westermann R (2012) Temporally coherent real-time labeling of dynamic scenes. In: 3rd international conference on computing for geospatial research and applications. ACM, pp 17:1–17:10

Wagner F, Wolff A, Kapoor V, Strijk T (2001) Three rules suffice for good label placement. Algorithm Special Issue GIS 2000:334–349

A Framework for the Automatic Geometric Repair of CityGML Models

Junqiao Zhao, Jantien Stoter and Hugo Ledoux

Abstract Three-dimensional (3D) city models based on the OGC CityGML standard have become increasingly available in the past few years. Although GIS applications call for standardized and geometric-topological rigorous 3D models, many existing visually convincing 3D city datasets show weak or invalid geometry. These defects prohibit the downstream applications of such models. As a result, intensive manual work of model repair has to be conducted which is complex and labour-intensive. Although model repair is already a popular research topic for CAD models and is becoming important in GIS, existing research either focuses on certain defects or on a particular geometric primitive. Therefore a framework that explores the full set of validation requirements and provides ways to repair a CityGML model according to these requirements is needed and proposed in this paper. First, the validity criterion of CityGML geometric model is defined, which guarantees both the rigorous geometry for analytical use and the flexible representation of geographic features. Then, a recursive repair framework aiming at obtaining a valid CityGML geometric model is presented. The geometric terms adopted in this paper are compliant with the ISO19107 standard. Future work will further implement the framework.

Keywords 3D models · CityGML · Validity · Repair

J. Zhao (✉) · J. Stoter · H. Ledoux
Department of GIS Technology, Delft University of Technology,
Delft, The Netherlands
e-mail: johnzjq@gmail.com

J. Zhao
Center for Spatial Information Science and Sustainable Development,
Tongji University, Shanghai, People's Republic of China

1 Introduction

Three-dimensional (3D) Geo-information has been treated as one of the essential sources of the latest and future GIS applications (Gruen 2008; Gröger et al. 2012; Stoter et al. 2013). However, current practice of digital city modeling focuses mainly on the visual appearance of a model rather than on its correctness of geometric-topological structure (Gröger and Plümer 2009). As a result, most of the analytical applications cannot be conducted with these ill-posed models, such as geometric processing, structural and environmental analysis and indoor navigation (Hughes et al. 2005; Isikdag et al. 2008; Haala et al. 2011). This is seen as a significant bottleneck of the 3D GIS industry and a waste of the expensive modeling efforts and expenses.

One of the error sources of 3D city models starts in the early stage of modeling. To produce visually convincing 3D city models with the least effort, producers often employ interactive modeling tools, e.g. 3D studio MAX, Maya, SketchUp and AutoCAD, to shape the appearance of a city objects with polygonal meshes. However, the freedom granted by these tools leads to imperfect mesh models (Botsch et al. 2007). In many projects, model templates are widely used which dramatically decrease the modeling costs for similar features among 3D city models (Badler and Glassner 1997). However, mismatches between model parts create various types of errors such as intersection (Fig. 1 *left*), overlap and gaps. Another error source is the model optimization. It is accomplished by vertex welding and simplification where close vertices and trivial triangles are merged. In this process, complex vertices and edges as well as degeneracies are often produced (Fig. 1 *right*). Moreover, modelers usually delete the concealed surfaces to further decrease the data size, which leads to the incompleteness of the model (Zhao et al. 2012).

Besides the errors that originate from the modeling process, errors can be created from conversion processing (Nagel et al. 2009). Many of the currently
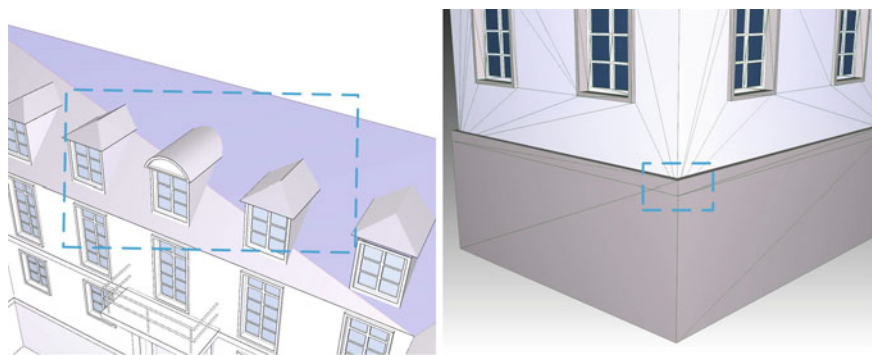
Fig. 1 Errors found in visually convincing 3D models (*left*, intersection between geometric parts; *right*, complex edges produced by model optimization)

available CityGML building models are produced from conversion of CAD models. The different definitions of building structures in the GIS and the CAD domain lead to inappropriate interpretation of building components. For example, walls are often modeled as *solids* in the CAD domain by applying the Industry Foundation Class standard (IFC), but they should be represented by *surfaces* in CityGML (Liebich et al. 2010; Gröger et al. 2012). For the sake of simplicity, the solid wall is always broken up into a set of *MultiSurfaces* during conversion, which omits not only the volume information of the component itself but also the topological relationships between walls. Additionally, errors may occur in the semantics editing process of CityGML models, if mistakes are made when manually matching geometries with CityGML semantics (Benner et al. 2005; Wagner et al. 2012).

Repair of 3D geometric models has become an important topic in the field of CAD and computer graphics. A number of repair methods have been proposed for polygonal meshes (Ju 2009; Campen et al. 2012). These approaches are however not sufficient for city models, since 3D city models are usually aggregated models, made up by heterogeneous geometric primitives in multiple dimensions. In the field of GIS, previous researches studied the validation of geometric models. Van Oosterom et al. (2004) proposed an extended definition and a set of validation rules for *polygons*. Kazar et al. (2008) introduced the validation rules for geometry defined in Oracle, especially the *surface* and the *solid*, but no repair method is provided. Verbree and Si (2008) and Ledoux et al. (2009) proposed methods to validate the GML *solids* based on tetrahedralizations. Gröger and Plümer (2012a, b) proposed rules and axioms for consistent representation and updating of 3D city models. However, their definition of geometry is not fully consistent with the ISO19107 standard, which acts as the geometric base of CityGML (Gröger et al. 2012).

Recently, a repair pipeline for CityGML models has been proposed in (Bogdahn and Coors 2010; Wagner et al. 2012). Although both the geometric and semantic aspects are included, they only repair errors that are similar to mesh repair, such as *holes* and incorrect *orientations* of mesh surface, while the overview of the different types of geometric errors of CityGML model concerning the downstream applications and their repair is not provided.

Starting with the definition of valid and invalid geometric model for CityGML in Sect. 2, a framework for the repair of geometric errors in a CityGML model is proposed in Sect. 3, which is a recursive framework and deals with the hierarchies of the CityGML model. Finally, we conclude with a discussion on the implementation issues and on future realizations.

2 Validity of CityGML Models

Before repairing, we should first agree on an exact definition of valid and invalid 3D city models. The ISO19107 standard provides the criteria of valid geometry, such as *simple* and *orientable* (Herring 2005). The question is whether these criteria are sufficient for the downstream applications.
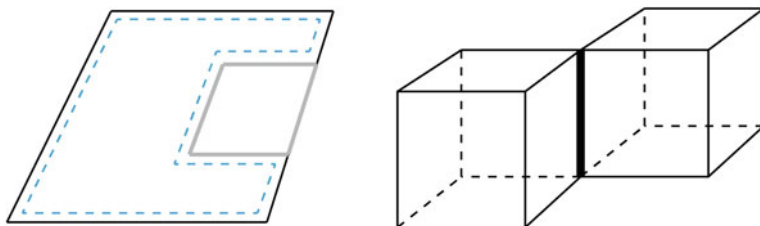
Fig. 2 Non-2-manifold *simple geometric objects* (*left*, a *simple surface* with its interior *ring* tangent to its exterior *ring*; *right*, a *simple composite solid*)

In visualization applications, such as urban planning and virtual reality, only the geometric primitives, i.e. *point*, *curve* and *surface* are demanded, thus no strict requirements is needed for the input geometry (Hearn and Baker 1996). In more sophisticated GIS applications, accurate and faithful representation of geographic features and their embeddings in the space are mandatory, such as 3D cadastre, thus the n-D *geometric object* should represents the continuous image of an n-D *space* (Herring 2005). This is defined as *simple* in the ISO19107 standard.

However, analytical applications demand stricter local and global properties for the input model. If restricted to the boundary-based representation, an *orientable* 2-manifold *surface*[1] (2-manifold for short) is always required (Botsch et al. 2007). However, the *surface* of a *simple geometric object* might not be a 2-manifold. Figure 2 *left* shows a *surface* with its interior *ring* (shown by the grey line) being tangent to its exterior *ring* (shown by the dark line) at the shared boundary edge. Because its *interior* (bounded by the dashed line in the figure) is isotropic, this *surface* is *simple* according to the definition (Herring 2005), but the shared *boundary* forms a dangling edge which is non-2-manifold. For a *simple* 3D *geometric object*, its *surface* can also be non-2-manifold because of the tangent edge (Fig. 2 *right*).

In conclusion, *simple* is not a sufficient validity criterion for the input model of all the downstream applications. The least requirement to support the geometric processing intensive applications is that all the input geometry should be 2-manifold. However enforcing 2-manifold globally for the representation of a collection of geographic features is intricate and difficult to model (Gröger and Plümer 2009; Thompson and van Oosterom 2011). And it is not able to represent the connection relationship between objects topologically without allowing tangency (Cavalcanti et al. 1997). Therefore, the validity of CityGML model should be defined in hierarchies according to the application requirements.

A CityGML model is composed of a hierarchy of features represented by instances of subclasses of *CityObject* (Gröger et al. 2012), we define the units of this hierarchy that to be used for analytical purpose as *component models*. Their

---

[1] a manifold of dimension *n* is a topological space in which each point has a neighbourhood that is homeomorphic to the Euclidean space of dimension *n*.

parents, collections of *component models* used for representation purpose, are termed as *aggregate models*. In practice, the classification of two types of models is customizable according to the practical requirements. Based on this definition, exact definitions of valid geometry of *component model* as well as of *aggregate model* are proposed in the next section.

2.1 Valid Geometry of Component Models

Based on the previous definition, the *component model* is the unit for geometric processing applications, thus 2-manifold should be the validity criterion for its geometry. If a *component model* is represented by only the 0D and 1D *geometric objects*, i.e. *point* and *line string*, *simple* is the sufficient validity criterion because *points* and *line strings* are the lower dimensional primitives of a 2-manifold. However, if 0D or 1D *geometric primitives* are presented in a higher dimensional (2D or 3D) *component model*, they should be eliminated because they form dangling cases which are non-2-manifold.

2.1.1 Validity Criteria for 2D Geometry

For a valid 2D *component model*, it should contain *surface* primitives with consistent orientations (the *surface* type is limited to planar polygon in CityGML). A non-closed *simple surface* without *holes* is a 2-manifold with boundaries,[2] thus is valid. The *closed simple surface*, a *shell*, represents a closed 2-manifold which is also valid. For a *simple composite surface*, all of the *surface*s included should be *simple*, and only touch at the existing shared *boundaries*, i.e. edges or vertices, as shown in Fig. 3a, b. Isolation and intersection of *surface*s shown in Fig. 3c and e are not allowed. The "free touching" of edges shown in Fig. 3d is also not allowed, because the combinatorial structure does not represent the shared primitives. Therefore, a *simple composite surface* without *holes* is also a 2-manifold. For a *simple MultiSurface*, the geometry should contain *simple surface*s with no intersections. However, the *surface*s can be isolated from each other or touch at an existing shared *boundary*. Both are valid cases that form the 2-manifold.[3]

If *holes* are present in a *surface*, its validity should be defined stricter than *simple* because of the possible cases of tangency existed in a *simple* geometry. If all the interior *ring*s of a *simple surface* are isolated and are located within the only exterior *ring*, the *simple surface* with *holes* is a valid 2-manifold with boundaries. However, if *ring*s touch, as shown in Fig. 4, non-2-manifold cases are formed, i.e. singular vertices (Fig. 4a), dangling edges (Fig. 4b, c). In Fig. 4d, the interior *ring*

---

2 Boundaries indicate the edges with only one neighbour face.
3 The disjoint union of a family of *n*-manifolds is a *n*-manifold (Lee 2010).
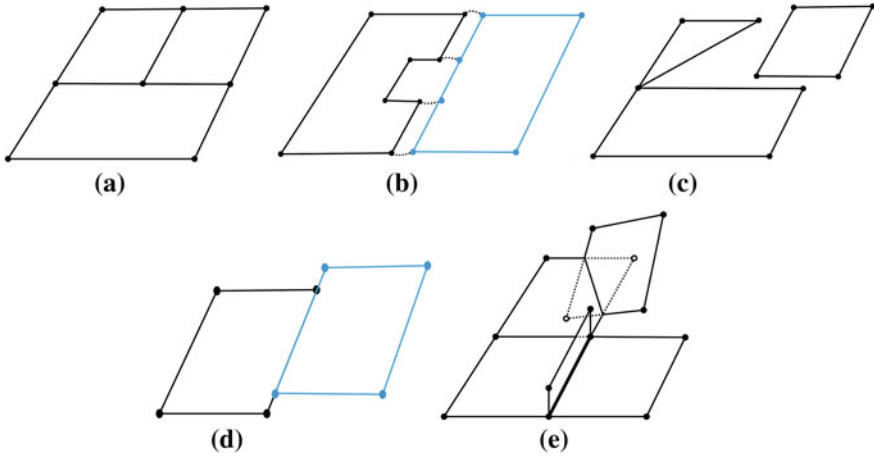
Fig. 3 Valid and invalid cases of a *composite surface*. **a** A valid *composite surface* which is also simple; **b** an valid *composite surface* where the *dashed line* indicates two identical vertices; **c** an invalid *composite surface* with one of its *surface* isolated; **d** an invalid *composite surface* containing a "free touching" between two *surfaces*; **e** an invalid *composite surface* with its *surfaces* intersected and forming a complex edge

breaks the interior of the *surface* into more than one connected component which is neither *simple* nor 2-manifold. The "free touching" situation shown in Fig. 4e is not reflected by the combinatorial structure, thus is also invalid. If *holes* are present in a *shell*, the *shell* is invalid according to its definition even it is a non-closed 2-manifold. For a valid *composite surface* with *holes*, the *surfaces* included should all be valid, and they should not intersect or disjoint, but touch only at shared existing boundaries in an appropriate manner. For a valid *MultiSurface* that contains *holes*, the only difference with the valid *composite surface* is that isolation of valid *surfaces* is allowed.

2.1.2 Validity Criteria for 3D Geometry

A 3D *component model* is represented by the 3D *geometric object*, i.e. *solid, composite solid* and *MultiSolid*. The validity requirements of 2-manifold should be
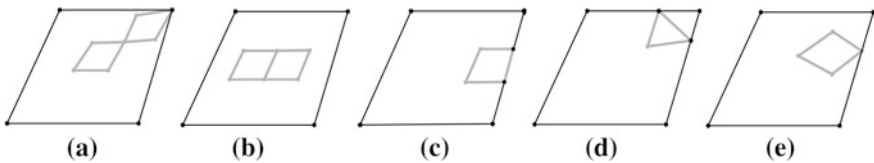
Fig. 4 Invalid *surface* with *holes* where vertices and edges of the exterior *ring* are colored in *black* while vertices and edges of the interior *ring* are colored in *grey*

verified on the *shells* of *solid* models. For a valid 3D *solid* without *voids* (*holes* in 3D), its exterior *shell* should be a closed valid *surface*, i.e. a compact, oriented 2-manifold without boundary. A 3D *solid* with a *handle* as shown in Fig. 5a can be treated as *simple* when referring to the connectivity of its *interior* and be used for representation purpose (Kazar et al. 2008). However, non-2-manifold situations are formed in the edges shown by bold lines. Figure 5b shows another case of *simple solid* with a non-2-manifold situation.

For a *composite solid*, even though each of the contained *solids* is individually valid, the touch between *solids* leads to complex edges, which form non-2-manifold cases (as shown in Fig. 5c by bold lines). As a result, the *composite solid* is not valid to represent a *component model* for analytical purpose. However, a *MultiSolid* with all its *solids* valid and isolated is valid.

If *voids* are presented in 3D geometry, interior connected *simple solids* as shown in Fig. 6b, c present non-2-manifold *complex edges* when two *shells* touch. As a result, a valid *solid* with *voids* should exclude any case of tangency between *shells*. Similar to the previous definition, a *composite solid* with *voids* is also not valid. And a valid *MultiSolid* with *voids* should contain only the isolated valid *solids* with *voids*, as shown in Fig. 6a.

A 3D component model can also contain non-closed 2D surfaces. However, the *surfaces* should be isolated from *solids* in order to avoid non-2-manifold situations via tangency.

2.2 Valid Geometry of Aggregate Models

An *aggregate model* is the assembling of *component models* used for representation purpose. The dimensions of *component models* can vary so the *aggregate model* is a heterogeneous set of *geometry objects* which can be represented by a *complex* or an *aggregate* as defined in ISO19107 (Herring 2005).

Since the valid *component models* are defined as 2-manifold, they can already be accepted by most of the analytical applications. To ensure the power of representation, non-2-manifold geometry can be allowed for an *aggregate model*. For
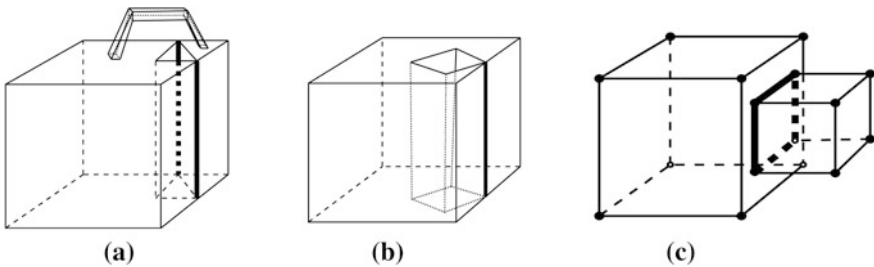
Fig. 5 Non-2-manifold cases formed in *simple solids* (**a**, **b**) and *composite solid* (**c**) with their *interior* connected
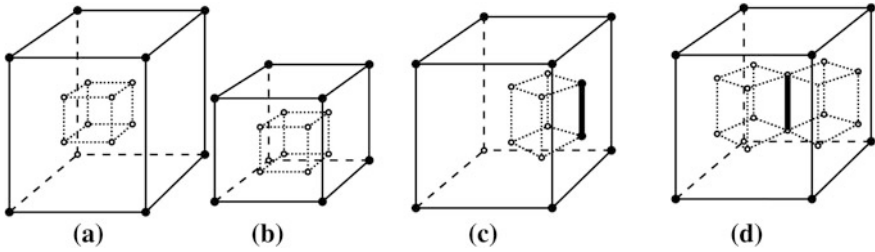
Fig. 6 Valid and invalid *solid* with *voids* **a** two valid *solid* with *void* (form a valid *MultiSolid*); **b** invalid *solid* with its interior *shell* touching the exterior *shell*; **c** invalid *solid* with its two interior *shells* touching

example, a building with an attached garage can be concisely represented by a *composite solid* as shown in Fig. 7 *left*, even though it is not a 2-manifold as a w*hole*. Otherwise, each of the buildings should be represented by a *solid* and certain relationships should be defined to describe their connection via walls. In another case, when representing two connected rooms inside a building (Fig. 7 *right*), it is proper to represent the model by a *complex*, which employs a *composite solid* to represent two connected rooms, and employs a *shell* to represent the exterior wall (Gröger et al. 2012). Because *aggregate* is a weak geometric form according to the standard, since intersection between *geometric objects* is allowed. *Complex* should be the valid geometric type *for aggregate models*.

3 Repair Framework of CityGML Models

3.1 The Recursive Paradigm

According to the hierarchical structure of CityGML model, the repair of CityGML model should be carried out in different levels in depth-first order. It is
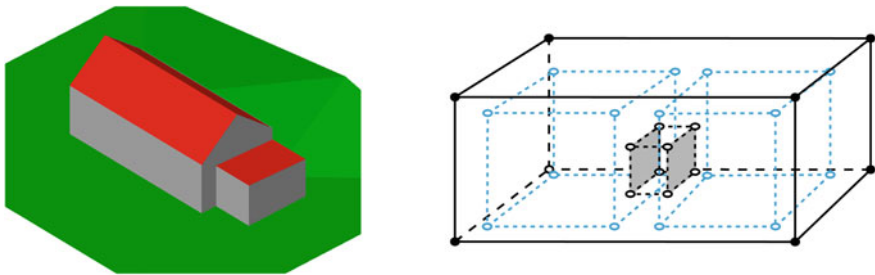
Fig. 7 Building represented by *complex*. **a** A building with a garage attached to one of its wall surface (Gröger et al. 2012); **b** demonstration of two connected rooms inside a building
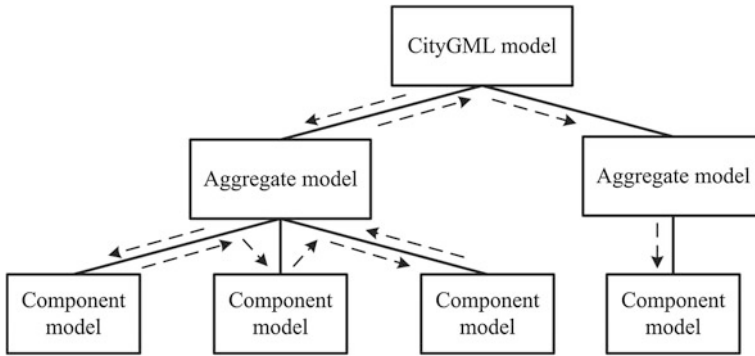
Fig. 8 The depth-first recursive framework for repair of CityGML models

straightforward to employ a recursive approach to fulfil the repair goal. After repairing all its *component models*, a higher level *aggregate model* is repaired in the next step (Fig. 8), until the w*hole* geometric model is valid as defined in Sect. 2.

3.2 Repair of the Component Model

The goal of this repair step is to convert the invalid geometry of a *component model* to 2-manifold. For the invalid 0D and 1D *geometric objects*, they have to be repaired to *simple* versions. This process includes the removal of duplicate *point*s (*P*1 and *P*2 shown in Fig. 9 left) by introducing a *tolerance* value, the conversion of the degenerated *line segments* ($\overrightarrow{P3P4}$ shown in Fig. 9 left) into *point*s, and the decomposition of the intersected *line segments* ($\overrightarrow{P3P7}$ and $\overrightarrow{P5P6}$ shown in Fig. 9 right). If a 1D *geometric object* is a *ring*, the closeness of the *line segments* should also be ensured.

3.2.1 Repair of 2D Geometry

For the 2D primitive, Ledoux et al. (2012) have proposed a triangulation based repair method for planar *polygons*. For a general *surface* embedded in a 3D

Fig. 9 Illustration of the repair of 0D and 1D geometry
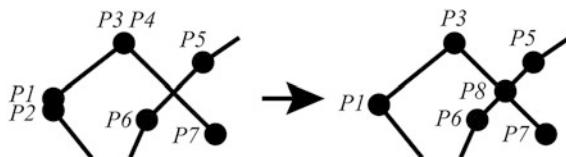
Table 1 Errors and their repair methods for 2D *surface*

Errors	Repair
Errors of 1D *rings*, non-planar (if it is a planar *polygon*)	The repair of *rings*
Degeneration	The removal of degeneracies by converting them to proper representations
Intersection and folding	The decomposition of the intersected *surfaces*
Complex edge	Cutting and stitching complex primitives
Invalid holes	Preventing touching of *rings* or decomposing the intersected *rings*
Dangling primitives	The removal of dangling edges and isolated vertices
Open shell	Hole filling for a *shell*
Isolation	Separating isolated *surface* parts
Inconsistent orientations	Restoring orientations

Euclidean space, repair for each of the possible error types summarized in Table 1 should be conducted.

First, *rings* of each *surface patch* should be validated and repaired using the repair methods for *line strings*. Besides, the planarity of each *ring* should also be ensured for *polygons* by introducing a *tolerance* value (van Oosterom et al. 2004). Then, the possible degeneration of *surface patches* should be checked. If degenerated, a *surface* should be converted to the proper primitive, i.e. a *line string* or a *point*. If a *surface patch* intersects with itself or with others in a set of *surface patches* (Fig. 10a), the intersection should be detected and the intersected *surface patches* should be decomposed by inserting vertices and edges (Fig. 10b). During this process, complex edges and vertices may appear. Algorithms cutting along these complex primitives and stitching them in an appropriate way should be used (Fig. 10c) (Guéziec et al. 2001). If *holes* touch, the touching boundaries should be offset in opposite directions. If a *hole* intersects with the exterior *ring*, the exterior *ring* should be decomposed as shown in Fig. 11. If an edge or a vertex has no neighbour faces, it is a dangling primitive and therefore should be removed

(a)   **(b)**   **(c)**
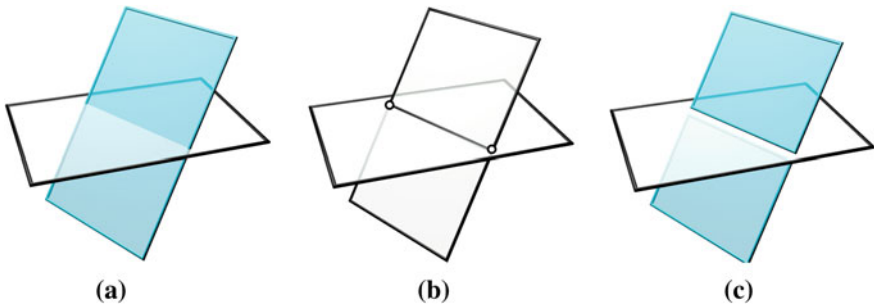
Fig. 10 Illustration of the repair of intersected 2D geometry **a** two intersecting *surfaces*; **b** decomposed result of (**a**); **c** the complex edge is splitted to avoid touching
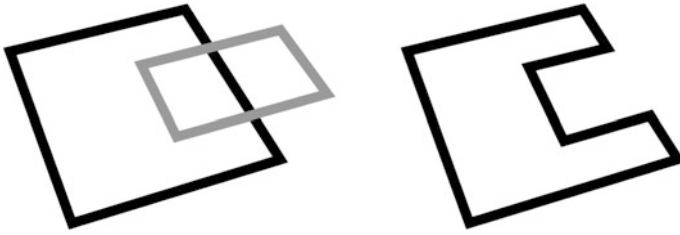
Fig. 11 Illustration of the repair of a invalid hole in a 2D *surface* where the *gray ring* indicates the *hole*
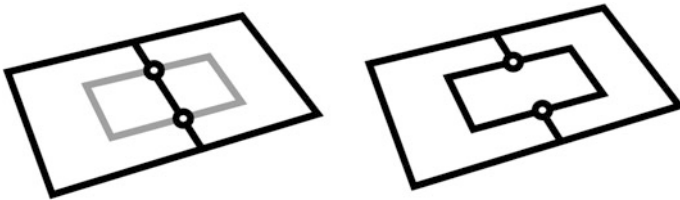
Fig. 12 Illustration of the repair of a dangling primitive in a 2D *surface* where the *gray ring* indicates a *hole*

(Fig. 12). If *holes* are present in a *shell*, *hole*-filling algorithms should be employed (Liepa 2003). If the *surface* contains several isolated parts, it should be separated into different *surface*s or be converted to a *MultiSurface*, which has to be repaired afterwards. Finally, the orientations of all the *surface patches* should be checked for consistency. If incorrect, the orientations should be restored by propagating from the orientation of a valid *surface patch*.

The repair of a *composite surface* is similar to the repair of a *surface* where *surface*s inside the *composite surface* can be treated as *surface patches* of a *surface*. If there are isolated *surface*s, the *composite surface* should be separated. For a *MultiSurface*, the repair pipeline of *composite surface* should be employed for each of the connected components.

3.2.2 Repair of 3D Geometry

For a 3D *solid,* all its *shells* should first be repaired based on the previous repair steps for the *surface*. Sometimes a *solid* may degenerate to a lower dimensional non-*solid* primitive, such as a *surface* or a *line string* or a *point*. The representation should be corrected accordingly. If there are *solids* that intersect with each other, the *solids* should either be decomposed and be converted to a *MultiSolid* or they should be merged into one *solid* if they share the same properties as shown in Fig. 13. Because the touching of *shells* produces complex cases, *solids* should be offset to avoid touching (Fig. 14). As discussed in Sect. 2.1, *voids* inside a *solid*
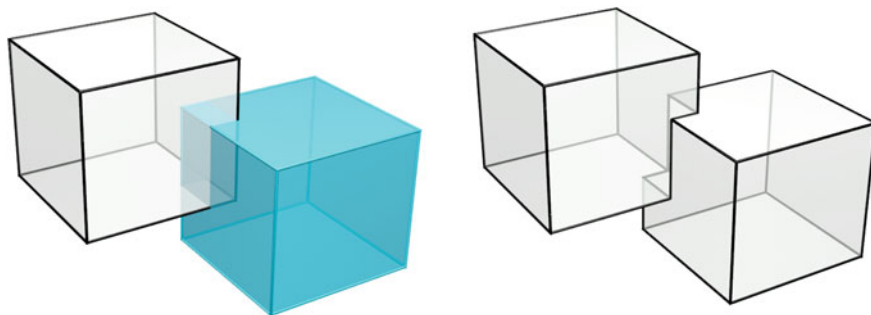
Fig. 13 Illustration of the repair of intersecting 3D *solids*

should not tangent to or intersect with the exterior *shell*. If this occurs, it should be healed by offsetting the tangent primitive inward the exterior *shell* or decompose the *exterior shell* as shown in Fig. 15. If lower dimensional primitives are presented, only the *surface* primitive should be repaired and kept (Fig. 16), while the 1D and 0D primitives should all be removed.

According to the validity criteria, a *composite solid* should be converted to a *MultiSolid* or be merged to a *solid*. Then the remaining separated *solid* can be repaired using the above steps which are summarised in Table 2.
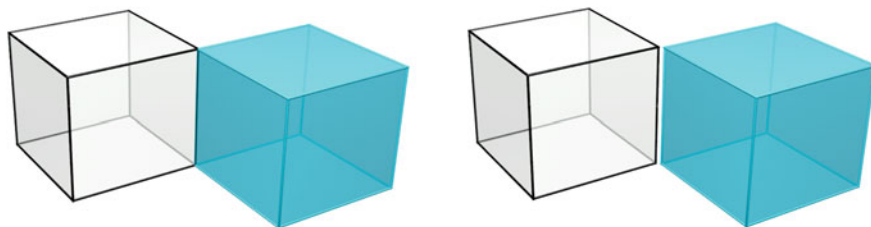
Fig. 14 Illustration of the repair of complex edges in a 3D geometry
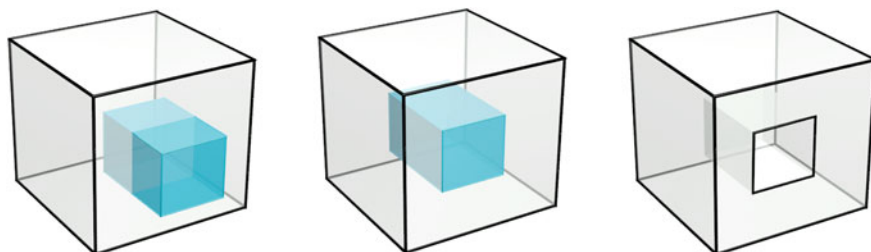
Fig. 15 Illustration of the repair of invalid *voids* in a 3D *solid* **a** the exterior *shell* and the interior *shell* intersect; **b** the exterior *shell* and the interior *shell* touch; **c** a repair result of both (**a**, **b**)
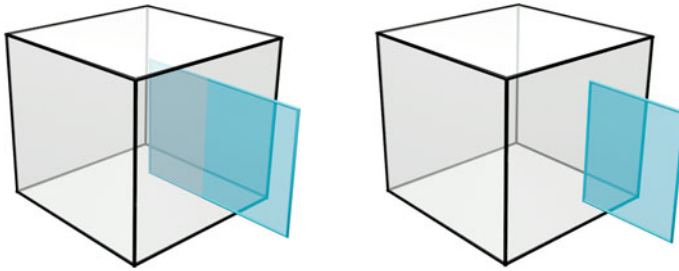
Fig. 16 Illustration of the repair of a lower dimensional *surface* in a 3D geometry

Table 2 Errors and their repair methods for 3D *solid*

Errors	Repair
Errors of 2D *shells*	The repair of *shells*
Degeneration	The removal of degeneracies by converting them to proper representations
Intersection	The decomposition or the merge of the intersected *solids*
Complex edge	Offseting the tangent *solids*
Invalid *voids*	Offseting tangent edges of *voids*
Dangling primitives	The removal of dangling edges and isolated vertices
Isolation	Separating isolated *solids*

3.3 Repair of the Aggregate Model

As discussed in Sect. 2.2, the valid geometry of the *aggregate model* should be the type of *complex*. The way to construct a *complex* from a collection of valid *component models* is straightforward. The ISO19107 standard had already provided a work flow (Herring 2005):

(a) If two primitives overlap, then subdivide them; eliminate repetitions until there is no overlap.
(b) Similarly, if a primitive is not simple, subdivide it where it intersects itself; eliminate repetitions until there is no overlap.
(c) If a primitive is not a *point*, calculate its *boundary* as a collection of other primitives, using, if possible, those already in the generating set, and insert them into the *complex*.
(d) Repeat step "a" through "c" until no new primitive is required.

During this repair process, the decomposition of the intersecting *component models* should not introduce invalid geometry to the *component models*.

3.4 Implementation Issues

In order to implement the framework, some crucial aspects should be considered.

First, the data structure should be able to faithfully record the input model without missing important information. With respect to geometry, a generic boundary-based representation should be adopted which supports the geometric types defined in CityGML and is also able to represent invalid non-2-manifold edges and vertices. Tolerance values are important to define the condition of degeneracy and planarity, and should therefore be defined carefully. Due to the uncertainty caused by possible rounding error, an exact predict method or even the exact arithmetic should be used (Richard Shewchuk 1997). Then, basic repair routines should be developed, such as *constrained triangulation*, which convert the input model into a *simplical 2-complex* representation, which easies the repair process; *regularization*, which removes dangling 0D or 1D elements (Granados et al. 2003; Worboys and Duckham 2004), and *intersection detection* and *decomposition*, which extract the intersecting *geometric primitives* and decompose the model based on the extracted intersections. More mesh repair methods like *hole*-filling, non-2-manifold curing should also be incorporated (Campen et al. 2012). Finally, the developed repair method should be modularized and conducted in an optimised order so that the repair process will not introduce new artefacts.

4 Conclusion

This paper provides a definition of valid geometric model for CityGML concerning both the rigorous geometry for analytical purpose and the flexibility for representation purpose. Possible errors of both the *component models* and the *aggregate models* defined for CityGML models are summarized and their healing steps are introduced within a repair framework. The goal of the repair is to achieve a valid CityGML model that can be used for most downstream applications other than only for visualization or representation.

Future work will comprise the implementation of the automatic repair routines discussed before based on a proper data structure. This paper mainly focuses on the geometric aspect of repair. The semantics of the CityGML model should be further exploited to strengthen the repair capability on specified models, such as a building. Finally, other repair issues listed in the ISO19157 standard (ISO 2012) should be studied in the follow-up research work.

References

Badler NI, Glassner AS (1997) 3D object modeling. In: SIGGRAPH 97 introduction to computer graphics course notes

Benner J, Geiger A, Leinemann K (2005) Flexible generation of semantic 3D building models. In: International ISPRS/EuroSDR/DGPF-workshop on next generation 3D city models, EuroSDR, Bonn

Bogdahn J, Coors V (2010) Towards an automated Healing of 3D urban models. In: Kolbe T, König G, Claus N (eds) Proceedings of international conference on 3D geoinformation. International archives of photogrammetry, remote sensing and spatial information science, Vol XXXVIII-4/W15. Shaker Verlag GmbH, Aachen, pp 13–17

Botsch M et al. (2007) Geometric modeling based on polygonal meshes, ACM SIGGRAPH 2007 courses. ACM, New York

Campen M, Attene M, Kobbelt L (2012) A practical guide to polygon mesh repairing. In: Proceedings of the 2012 Eurographics, Cagliari, Italy, pp t4

Cavalcanti PR, Carvalho PCP, Martha LF (1997) Non-manifold modelling: an approach based on spatial subdivision. Comput Aided Des 29(3):209–220

Granados M et al. (2003) Boolean operations on 3D selective Nef complexes: data structure, algorithms, and implementation. In: Di Battista G, Zwick U (eds) Algorithms—ESA 2003: 11th annual european symposium. Lecture notes in computer science. Springer, Budapest, Sept 2003, pp 174–186

Gröger G, Plümer L (2009) How to achieve consistency for 3D city models. GeoInformatica 15:137–165

Gröger G, Plümer L (2012a) Provably correct and complete transaction rules for updating 3D city models. Geoinformatica 16(1):131–164

Gröger G, Plümer L (2012b) Transaction rules for updating surfaces in 3D GIS. ISPRS J Photogramm Remote Sens 69:134–145

Gröger G, Kolbe TH, Nagel C, Häfele K (2012) OpenGIS city geography markup language (CityGML) encoding standard version 2.0.0. Open Geospatial Consortium. OGC 12-019

Gruen A (2008) Reality-based generation of virtual environments for digital earth. Int J Digit Earth 1(1):88–106

Guéziec A, Taubin G, Lazarus F, Hom B (2001) Cutting and stitching: converting sets of polygons to manifold surfaces. IEEE Trans Visual Comput Graph 7(2):136–151

Haala N, Fritsch D, Peter M, Khosravani A (2011) Pedestrian navigation and modeling for indoor environments. In: Proceeding of 7th international symposium on mobile mapping technology, Cracow

Hearn D, Baker MP (1996) Computer graphics, C version. Prentice Hall, Englewood Cliffs, 652 pp

Herring JR (2005) ISO 19107:2005: geographic information-spatial schema. International Organization for Standardization

Hughes T, Cottrell JA, Bazilevs Y (2005) Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. Comput Methods Appl Mech Eng 194(39–41):4135–4195

Isikdag U, Underwood J, Aouad G (2008) An investigation into the applicability of building information models in geospatial environment in support of site selection and fire response management processes. Adv Eng Inform 22(4):504–519

ISO (2012) ISO 19157: Geographic information—data quality. International Organization for Standardization

Ju T (2009) Fixing geometric errors on polygonal models: a survey. J Comput Sci Technol 24(1):19–29

Kazar BM, Kothuri R, Oosterom P, Ravada S (2008) On valid and invalid three-dimensional geometries. In: Advances in 3D geoinformation Systems. Lecture notes in geoinformation and cartography. Springer, Berlin, pp 19–46

Ledoux H, Verbree E, Si H (2009) Geometric validation of GML solids with the constrained delaunay tetrahedralization. In: Proceedings of the 4th international workshop on 3D geo-Information, Ghent

Ledoux H, Ohori KA, Meijers M (2012) Automatically repairing invalid polygons with a constrained triangulation. In: Proceedings of Agile 2012, Avignon

Lee JM (2010) Introduction to topological manifolds. Graduate texts in mathematics, vol 202. Springer, New York, 433 pp

Liebich T et al. (2010) Industry foundation classes 2×4 (IFC2 $\times$ 4) release candidate 2

Liepa P (2003) Filling holes in meshes. In: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing, Aachen, Germany, pp 200–205

Nagel C, Stadler A, Kolbe TH (2009) Conceptual requirements for the automatic reconstruction of building information models from uninterpreted 3D models. In: Kolbe TH, Zhang H, Zlatanova S (eds) GeoWeb 2009 academic track. Cityscapes, Vancouver

Richard Shewchuk J (1997) Adaptive precision floating-point arithmetic and fast robust geometric predicates. Discret Comput Geom 18(3):305–363

Stoter J et al. (2013) Implementation of a national 3D standard: case of the Netherlands. In: Pouliot J, Daniel S, Hubert F, Zamyadi A (eds) Progress and new trends in 3D geoinformation sciences. Lecture notes in geoinformation and cartography. Springer, Berlin, pp 277–298

Thompson R, van Oosterom P (2011) Connectivity in the regular polytope representation. GeoInformatica 2(15):223–246

van Oosterom P, Quak W, Tijssen T (2004) About invalid, valid and clean polygons. In: Fisher PF (eds) Developments in spatial data handling. 11th international symposium on spatial data handling, pp 1–16

Verbree E, Si H (2008) Validation and storage of polyhedra through constrained delaunay tetrahedralization. In: Geographic information science. Lecture notes in computer science. Springer, Berlin, pp 354–369

Wagner D et al. (2012) Geometric-semantic consistency validation of CityGML Models. In: 3D geoInfo conference 2012, Quebec City, May 16–17

Worboys M, Duckham M (2004) GIS: a computing perspective. CRC, Boca Raton

Zhao J, Stoter J, Ledoux H, Zhu Q (2012) Repair and generalization of hand-made 3D building models. In: Proceedings of the 15th workshop of the ICA commission on generalisation and multiple representation jointly organised with EuroSDR commission 4—data specifications, Istanbul, p 10

Augmented Reality Visualization
of Archeological Data

Daniel Eggert, Dennis Hücker and Volker Paelke

Abstract One intention of archeology is the documentation and reconstruction of historical development of mankind. The extracted data of an archeological excavation is usually spatial referenced and visualized with the help of maps or geographical information system. Both, paper maps and digital representations have partly complementary strengths and shortcomings in their application. With Augmented Reality, both Systems can be combined and complement each other. This Work presents a concept for augmenting archeological paper maps with 3D models and additional interaction options. Besides the presentation of contents in 3D space for museum visitors, the identified examples of usage include the generation of new contents to support the archeological work on an excavation site. The mobile application *ARAC Maps* (Augmented Reality for Archeological Content) realizes this concept based on commercially available devices with the Android operation system.

Keywords Augmented reality · Archaeological mapping · 3D geovisualization · Google android

1 Introduction

Augmented Reality (AR) is a technology that extends the real environment with computer-generated content. The history of AR can be traced back to the development of the ultimate display, an AR system with a head-mounted display

D. Eggert (✉) · D. Hücker · V. Paelke
Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover, Hanover, Germany
e-mail: eggert@ikg.uni-hannover.de

developed by Sutherland (1968). Progress in computer technology development in the early 1990s led to an expansion of research and development in AR technology, but AR applications remained limited to laboratories or special installations for a long time, mostly due to the high performance hardware required. Mobile augmented reality was finally enabled on a large scale by the development of devices like smartphones and tablets that combine potent processing and graphics hardware with the necessary sensors.

With the help of AR visualizations information can be interactively communicated from a totally new perspective. Archaeology is one of the most promising applications that can be improved by AR presentations. It is the nature of archaeology, that the structures and objects (artifact's) uncovered by excavations are often incomplete or damaged. Reconstruction of past structures and artifacts is therefore an important task, especially when the results are to be communicated to a general audience. Typical presentation formats in museums range from posters and maps to actual physical reconstructions. However, physical models are costly and are therefore not always viable. More common are reconstructions presented as smaller scale models or the use of rendered 3D models in video presentations. While such models and videos can present relevant information more effectively than 2D depictions, they limit the transfer of knowledge because they are static and not interactive. Interactive 3D visualizations, and especially AR applications, can improve the communication process because they enable interaction and can also directly tie the presented information to the existing site or artifacts. AR systems have many potential benefits: due to the proliferation of high performance smartphones and tablets the required hardware is relatively cheap and many visitors can in fact use their own device. An additional aspect is the simplified navigation compared to purely virtual environments. Because the AR content is spatially linked to physical artifacts the users can employ their everyday navigation skills to explore the content and do not need to learn a new set of controls. Often interesting 3D content such as reconstructions are already available as a result of the archaeological work but only a small portion can be presented using conventional presentation formats. AR thus has high potential to make a larger selection of contents available, to dynamically adapt the presented content to user preferences and to enable user interaction with the presented information. The challenge is to make the implementation of such AR systems practical and inexpensive for the potential content creators, e.g. archaeologists or museum curators.

2 Related Work

Our work presented in this paper relates both to the augmentation of maps and the use of augmented reality in the presentation of archaeological content. While most AR systems aim to augment a physical environment, the augmentation of individual objects in general and paper maps in particular has been studied by several researchers in the past. An even larger application area has been the use of AR

techniques within an archaeological context. We can therefore only provide an overview of several projects that are of particular relevance to our system.

ARCHEOGUIDE (Augmented Reality based Cultural Heritage On-site GUIDE) was an EU-funded project to develop an AR system for visitors to an archaeological site (Vlahakis et al. 2002). The project was centered around the idea of providing visitor to an archaeological site with a mobile AR system, consisting of a portable computer, a HMD (head-mounted-display) equipped with a camera and a GPS sensor for positioning. Based on the interests and requirements of the user the system would then propose a route through the excavation site. During the site visit the system superimposes 3D models of reconstructed monuments into the users view, using the HMD, shown in Fig. 1. Users can interact with the presented content through pointing gestures, e.g. to obtain additional audio information on an object of interest. While ARCHEOGUIDE used custom hardware like head-mounted-displays, later projects have also investigated the use of off-the-shelve hardware in similar application settings. The work on AR in archaeology has focused on the presentation of information to museum visitors. The use of AR as a tool to support archaeologists in the field has remained largely unexplored.

While systems like ARCHEOGUIDE are designed to work in the specific site of interest applications that augment artifacts like maps have the advantage that they can be used in many sites (e.g. museums, classrooms) simultaneously. One of the first systems for augmenting paper maps with the AR technologies was developed by Bobrich and Otto (2002). Their system used an optical tracking system with specific marker patterns that were attached to the paper map to determine the position of the user with respect to the map. The display of the 3D augmentation information was realized with a HMD. For interaction the user had the opportunity to use additional markers within the field of view of the camera in order to trigger predefined actions, as shown in Fig. 2. As the application depended on a desktop PC for the tracking and rendering the mobility of the user was limited.

Mobile systems for augmenting paper maps were presented by Reilly et al. (2006) for a system that used RFID chips for tracking and Schöning et al. (2006)

Fig. 1 Example of on site augmentation technique, generated by ARCHEOGUIDE
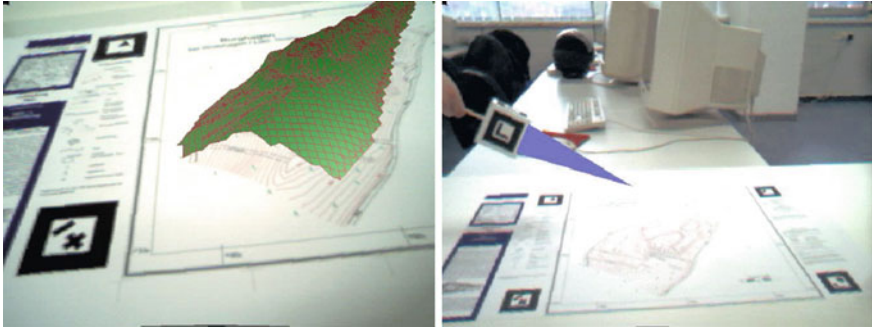
Fig. 2 Representation of a terrain model with augmented maps (Bobrich and Otto 2002)

for a system using marker patterns. Another augmented paper map system, augmenting charts for pleasure crafts was described by Paelke and Sester (2010), using microscopic dot patterns for tracking. Schmalstieg and Wagner (2006) describe a mobile, marker-based application for virtual scavenger hunts within museums. Many of these early works were largely technology demonstrators due to the technological limitations and the requirement for expensive specialized hardware like HMDs. Another limitation was the need to integrate additional infrastructure (like RFID chips) or information (marker patterns) into the map to enable the tracking functionality.

3 Concept and Requirement Analysis

In order to determine the requirements of the prototype, we interviewed archeologists from the *Institut Català d'Arqueologia Clàssica*. We identified two user groups, the museums visitors and archeologists working on an excavation site, also referred to as experts. For the user group of visitors, we are focusing on communicating information in an intuitive way. This includes the visualization of 3D models and the delivery of textual and visual information to support the knowledge transfer. The requirements of the experts focus on the generation of new data. We have implemented four potential use cases, two for each user group and integrated them into the archeological excavation process.

3.1 Excavation Process

The typical archeological excavation process, as shown in Fig. 3, consists of four main components. This process is by nature destructive and needs accurate documentation to reconstruct the excavated structures later on. The tools of choice for this documentation and presentation are maps, in particular paper maps.
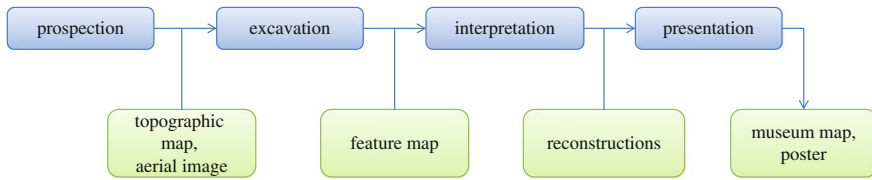
Fig. 3 Archeological excavation process and its products

The prospection of an area with techniques like airborne photogrammetry or geomagnetic surveys leads to maps of potentially interesting excavation sites. During the excavation operations soil layers are removed to uncover the ancient structures beneath. The positions of structures and artifacts found are mapped before the removal from the site. The interpretation of structures and finds leads to reconstructions of buildings and other spatial referenced information, e.g. trade relations between empires. The final step in the excavation process is the presentation of all the data gathered in the previous steps. The data is usually presented in 2D on paper maps or posters.

The two use cases for the user group of visitors fit into the presentation phase of the excavation process. The data collected in the previous step, like reconstructed buildings should be visualized in 3D on the paper maps. Besides the 3D visualization, the knowledge acquisition should be supported by additional information of the visualized models.

For the user group of experts, we focused the generation of new data to support the field work. First, we will implement an annotation function. This will aid the documentation during the destructive and irreversible excavation phase. The second use case will cover the field of generative models. The phase of excavation and interpretation is usually an iterative process driven by assumptions that will need to be proofed through digs at the excavation site. The possibility to create models based on assumptions directly at the excavation site and manipulate them according to new finds can combine these steps in field and speed up the whole process significantly.

3.2 Comparison of Paper and Digital Maps

The shown archeological workflow includes the use of paper maps. While the relationship between paper and digital maps are often denoted as competing, Dymetman and Copperman (1998) describe it as complementary. This derives from their properties. The content of a paper map is fixed and unchangeable, therefore they are considered permanent. Furthermore they are cheap, lightweight and hence easy to transport. Provided in a particular size a paper map may be used by multiple users.

In contrast digital maps are dynamic and can be adapted to various applications. They provide interaction capabilities and may extent the visualization displaying 3

dimensional models. The major problem of digital maps is need for a rather expensive infrastructure. This results in restricted mobility and is susceptible to technical errors.

Still, the discussed archeological use cases will benefit from the advantages provided by digital maps. Considering the museum use case, a map has to provide (additional) information and must be fairly cheap. Paper maps are cheap but the provided information is predefined and fixed. In terms of communication theory this is called a *push* medium. A proper information provider, however, also includes *pull* capabilities. Unlike paper maps, digital maps may provide such a feature, where users can interact with the map and acquire additional information of interesting areas. The distribution of information is thus much more personalized, efficient and attractive.

Moreover, digital maps are already widely used in the archeology domain. Computer-aided design (CAD) and GIS are important components of the archeological workflow. Because of their easy handling, paper maps are nonetheless the main map-tool at an excavation site.

4 Implementation

The goal was to develop a practical and cost effective platform for AR applications that work on common smartphones and tablets, thereby making them accessible to large audiences. Due to the advances of mobile hardware in processing power and the integration of custom hardware for 3D rendering we were able to use natural features tracking, avoiding the need to integrate special marker patterns for tracking.

Our prototype is implemented on Android OS and thus runs on a wide range of mobile devices like smartphones and tablets. The image matching to recognize the paper maps as targets for the AR content is based on the Vuforia SDK. This free-to-use toolkit features a fast and reliable image tracking and is compatible with most Android devices. It supports the generation of new image markers with an online web-tool. In case a predefined marker is recognized in the video frame, the pose of the camera relative to the marker is estimated from the distortion of features found. This pose is used to render the 3D scene and display it on the marker.

The prototype supports VRML and PLY-format files for rendering triangulated 3D models as well as simple point sets, as shown in Fig. 4. In order to combine the models with the markers, a central configuration file with XML layout is used, defining a 3D scene on the markers with different model positions and scales.

The prototype features four different functions to support use cases in the fieldwork and presentation phase of the excavation process. The presentation of 3D models and additional information is targeting the user-group of museum visitors with the focus on knowledge transfer. The functions to generate a street grid or place digital annotations on the paper map targets the group of archeologists and focus on generating new data to support the field-work.
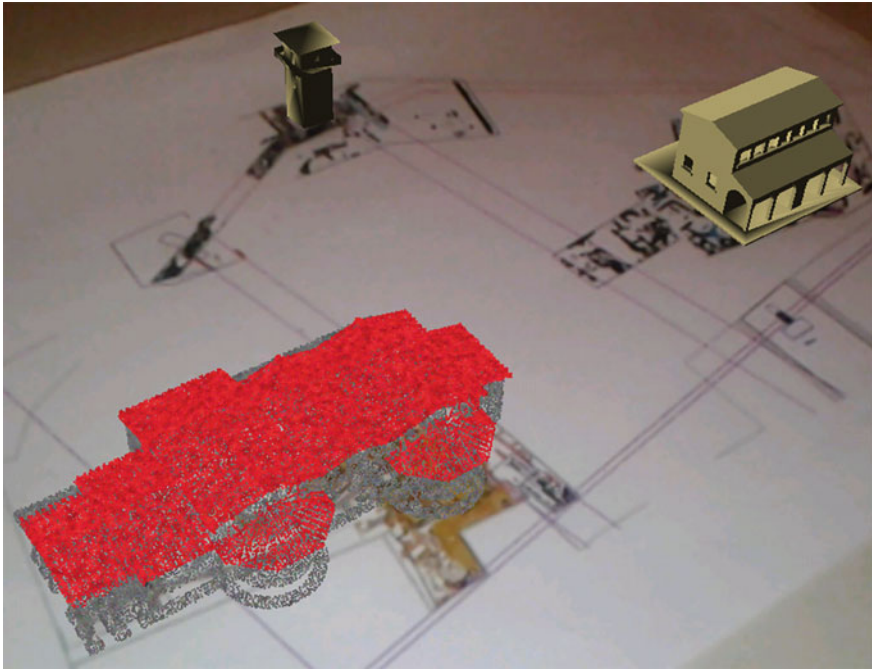
Fig. 4 Archeological paper map augmented with simple point sets and triangulated models

4.1 Layers

The prototype supports VRML and PLY-format files for rendering triangulated 3D models or simple point sets. To organize the models, a layer system similar to a GIS is used. The layers defined in a separate configuration file can hold an unlimited number of 3D models. Organizing models in layers, as shown in Fig. 5, lots of information can be presented easily on the paper map while the users won't lose track through too much input.
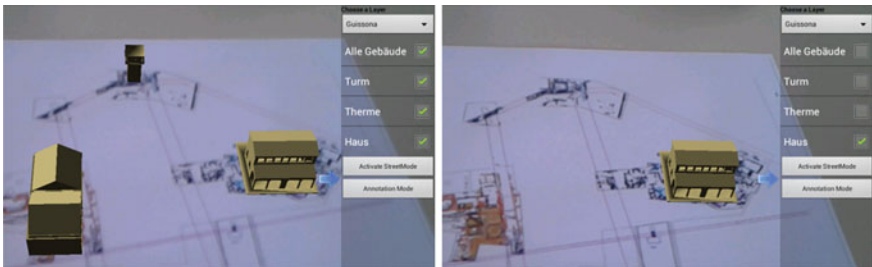
Fig. 5 Layer selection menu

4.2 Model Information

To display additional model information we use HTML files which can be saved locally on the device or as a global accessible web page. This way the information can be changed dynamically without the need of altering the actual application on each device. In case the user touches a model on the screen, the information is shown in a small HTML-based overlay depicted in Fig. 6.

In order to check if the user touched a model on the screen, the 2D touch position needs to be converted into the 3D scene. In computer graphics this task is usually referred to as picking. The usual approach is to transform the 2D touch-position into a 3D ray, and intersect it with every object. The computational effort of this intersection depends on the number and complexity of the models and is done fully by the central processing unit. This is rather unfavorable, because the tracking already needs most of the computational power. Therefore we implemented a color picking algorithm, where the scene is rendered in background with all models having a unique color, as shown in Fig. 7.

After reading the color at the touch position of this scene, the color code is dereference back to the corresponding model. The advantage of this approach is that the computation is mostly done by the graphics hardware, leaving the CPU for the tracking.
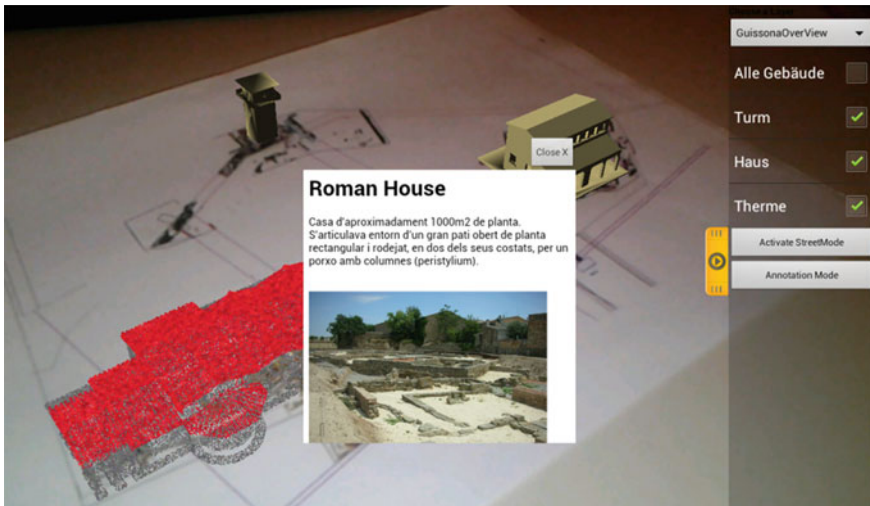
Fig. 6 Additional model info shown as HTML overlay
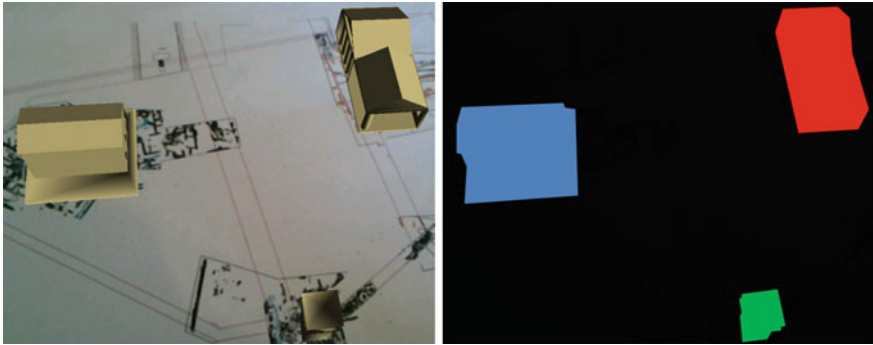
Fig. 7 Color picking with color coded models

4.3 Generative Street Models

A very useful find on a roman excavation site are ruins of ancient streets. Figure 8 shows the street model of a typical roman settlement as a well-organized street grid.

Interpreting the already found street structures gives the scientists hints where to find the structures of buildings in between the street grid and therefore another dig is worthwhile. We use this scenario to implement a real-time modifiable street network as an example for generative models in general.

The generation of a street network model consists of five steps depicted in Fig. 9. First, the user defines a polygon indicating the town boundary by touching the desired positions on the display. In a second step, the application calculates the center of the polygon. Third, the two perpendicular main streets called *decumanus* and *cardo* going through the center point are added to the street model. After-wards, the boundary's bounding box is filled with streets parallel to the main streets building the grid-based street network. Finally the street network is clipped to the initially created boundary.

After the initial generation, the street network's orientation, translation and grid size can be intuitively manipulated using common multitouch gestures as Fig. 10 illustrates.

With such an interactive street model, the archeologist can verify their assumptions and keep them up-to-date right at the excavation site without waiting for the office work.
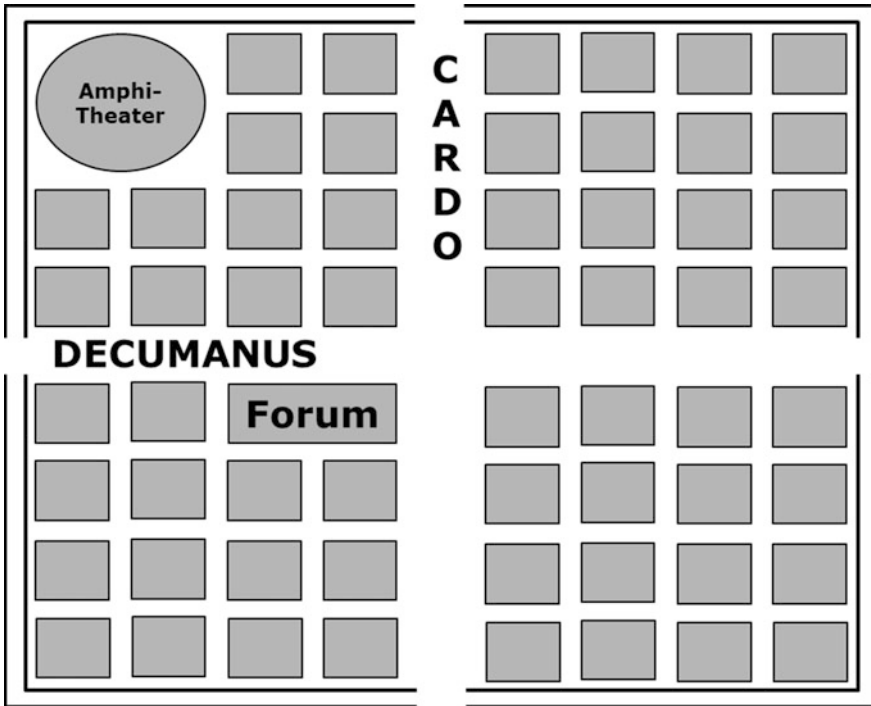
Fig. 8 Well-organized street grid of a roman settlement
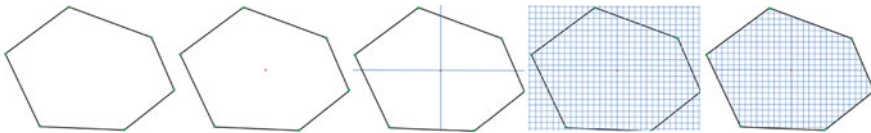
Fig. 9 Generation of a roman street network model

4.4 Annotations

Pen annotations on paper maps are a well-known application, but they lack some features which digital annotations can offer. First, a pen annotation changes the map content permanently and irreversible. Second, there is only a limited amount of space on a paper map. Yet, the further use of the notes in a spatial context, e.g. a GIS, is problematic, as it is not georeferenced. Digital notes can bypass these problems. The spatially located notes could be easily transferred to a geo-database for further use. At the current state, the application allows to add, display and delete an arbitrary amount of notes on the paper map area. As shown in Fig. 11, the individual notes are visualized as teal colored rectangles on the map.
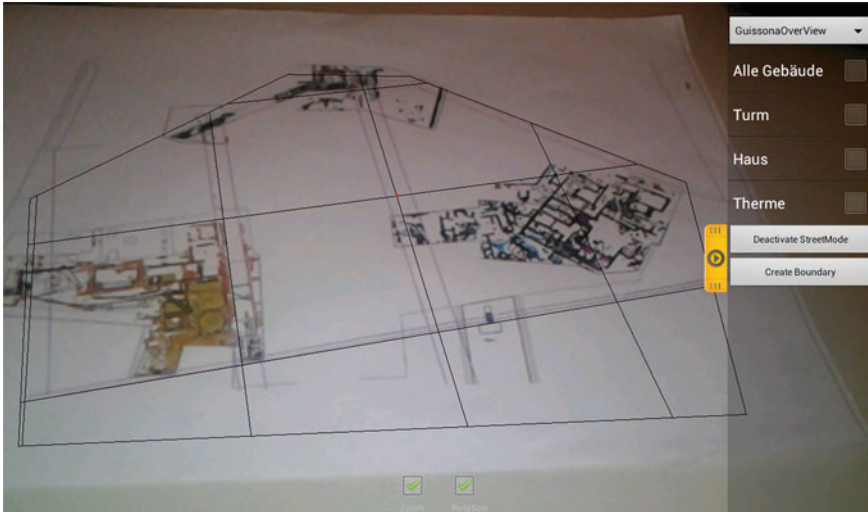
Fig. 10 Archeological paper map augmented with a street network model

5 Early Experiences and Feedback

The applications usability was tested with a small group of persons representing the non-experts or in other words the user group of visitors. (ISO-9241-11 1998) defines usability as "the effectiveness, efficiency and satisfaction with which specified users can achieve specified goals in particular environments". To test these specific usability criteria, we set up a test consisting of four tasks covering all implemented use cases. During the test we logged empirical analyzable data like the time or clicks required to accomplish the tasks. Besides this data, we checked the user's satisfaction with a questionnaire at the end of the test. The test setting ignores the user group of the experts and therefore can't give us hints whether the implemented expert functions can give a real advantage over already existing solutions, but we can identify potential problems in the workflow or the ease of use and learnability of the functions.

The analysis of the data revealed minor problems in the field of usability. For example, the users complained about the small and hard to discover menu handle on the right side of the application. As shown in Fig. 12, we changed this menu handle to a bigger one with more contrast to the underlying paper map, dominated by white and grey colors.

Another problem arose from the street network generation task. Aiming for expert users this task involves several consecutive actions which have to be carried out by the user in the correct order. The test revealed this task as rather error-prone for non-expert users. The test users often were not aware of the actions needed in order to create the town boundary and how to manipulate the street network in it.
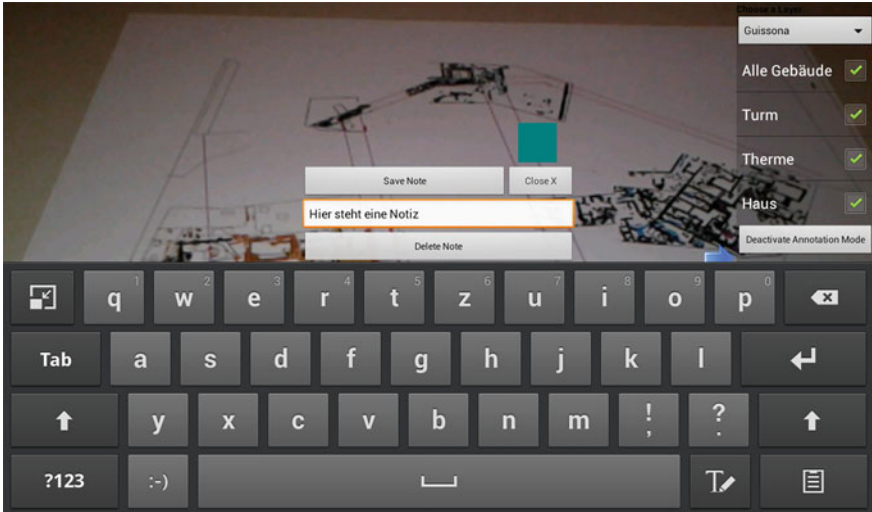
Fig. 11 Archeological paper map augmented with an annotation
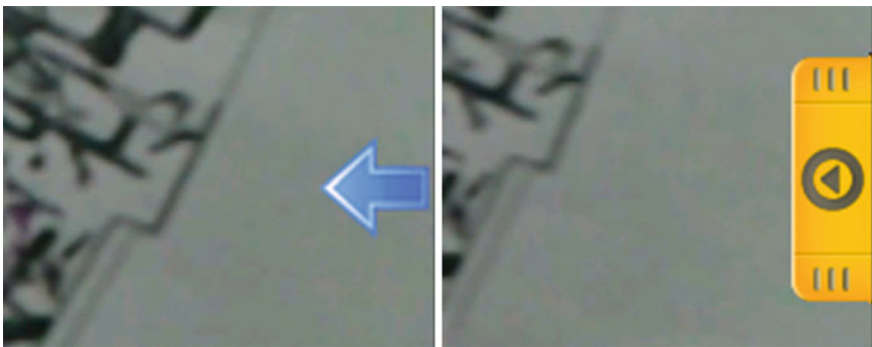
Fig. 12 Bigger menu handle with more contrast increased the usability

Thus, the app supports the user by showing tooltips popping in the screen guiding him through all necessary steps, as shown in Fig. 13.

The analysis of the questionnaire shows an overall positive view from the users. This indicates that the prototype is generally accepted by the users and the use cases, especially the knowledge transfer for the visitor's user group benefits from the AR visualization.
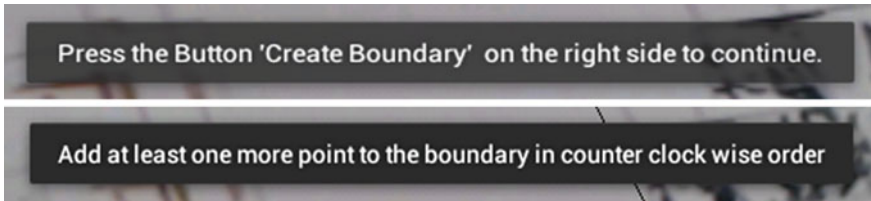
Fig. 13 Tooltips guiding the user through a task

6 Conclusions and Outlook

This research evaluated the applicability of Augmented Reality in the archeological domain. Potential requirements and corresponding use cases were identified in collaboration with domain experts from the *Institut Català d'Arqueologia Clàssica*. From those requirements we derived a concept for an AR-based Application, which we implemented in a prototype Android App. The core of the developed concept is the augmentation of paper maps with 3D interactive content.

The 3D visualization and the distribution of additional information are use cases focusing on non-expert users. The implemented interaction modes are the selection of augmented 3D models for additional information presentation and a layer grouping for these models. In contrast, the use cases focusing on expert users include more sophisticated interaction schemes. The developed prototype implements two expert user interaction schemes for generating additional content. First is an annotation mode, allowing expert users to place georeferenced annotations as augmentation on top of a paper map. Second is a street generation mode, enabling the dynamic generation and manipulation of a typical grid-based roman street-network. Both features do not alter the original paper map in any way.

The prototype implementing the described features is called *ARAC Maps* (*Augmented Reality for Archeological Content on Maps*) and bases on the Android platform. The conducted evaluation revealed some minor usability flaws. Addressing those problems improved the usability as well as the stability of the prototype.

Nonetheless, the implementation of all identified requirements and ideas was not possible within the context of this research. So our future work involves the support for more data formats other than VRML and PLY as well as texture and level-of-detail support. Furthermore, the findings of an excavation may originate from different time periods. This fact introduces a time layer feature, assigning a time component to each object or entire layers. With reconstructed 3D building models tagged with such a time component the development of historical sites can be easily and interactively comprehended. Moreover the augmentation of the excavation site itself represents a desired feature. This however, lies beyond the scope of augmenting paper maps with digital content. Finally a comprehensive evaluation with archeologists as the targeted expert user group has to be conducted.

References

Bobrich J, Otto S (2002) Augmented maps. In: Geospatial theory, processing and applications, Ottawa

Dymetman M, Copperman M (1998) Intelligent paper. In: Electronic publishing, artistic imaging, and digital typography, vol 1375, pp 392–406

ISO-9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs)—part 11: guidance on usability

Paelke V, Sester M (2010) Augmented paper maps: exploring the design space of a mixed reality system. ISPRS J Photogramm Remote Sens 65:256–265

Reilly D, Rodgers M, Argue R, Nunes M, Inkpen K, Lab E (2006) Marked-up maps: combining paper maps and electronic information resources. Pers Ubiquit Comput 10:215–226

Schmalstieg D, Wagner D (2006) A handheld augmented reality museum guide. In: IADIS international conference on mobile learning

Schöning J, Krüger A, Müller HJ (2006) Interaction of mobile camera devices with physical maps. In: Adjunct proceeding of the fourth international conference on pervasive computing, pp 121–124

Sutherland IE (1968) A head-mounted three dimensional display. In: Proceedings of the joint computer conference, part I, pp 757–764

Vlahakis V, Ioannidis N, Karigiannis J, Tsotros M, Gounaris M, Stricker D, Gleue T, Daehne P, Almeida L (2002) Archeoguide: an augmented reality guide for archaeological sites. IEEE Comput Graph Appl 22:52–60

The Virtual Centimeter World Model

Franz Leberl

Abstract Limitless sensing at ever greater detail, storage at nearly no cost and GPU-enhanced high performance computing have vastly reduced previous limits to the processing and use of digital images in Computer Vision. If citizens collect images of ever improved quality at a centimeter pixel size and with great density, thus high image overlaps, of our environment, if Internet-based image management systems assemble these photographs to meaningful image blocks at quantities in the realm of Exabytes, when 1 million images can be processed per day fully automatically into 3D Geo-information, can we then expect an emergence of very detailed 3D models of our entire urban and rural World? We argue that yes, 3D models of the World are feasible at a detail in the range of centimeters with current technology. Since that technology continues to evolve, the likelihood increases rapidly that such detailed World models will be created. Global aerial orthophotos in the decimeter range are being produced today; centimeter-type pixels are being collected along the entire street network of major cities. Very little is needed to convert such data into the reality of the Virtual Centimeter World Model at pixel-accuracy for a mixed reality experience.

Keywords Internet mapping · Web-based mapping · Neo-photogrammetry · 3D world model · Location-aware Internet · Global orthophoto · Street-side imagery · 3D urban models · Mixed reality

1 From Digital Mapsto Internet-Inspired Mapping

We now call the Internet "Location-Aware" (Leberl 2007; Leberl and Gruber 2008; 2009a, b) since it associates a Geo-position with many objects of an Internet search. Figure 1 illustrates a typical Geo-location on the Internet, in this case

F. Leberl (✉)
Institute for Computer Graphics and Vision, Graz University of Technology, Graz, Austria

Fig. 1 Virtual urban 3D model of Manhattan from BING/Maps (*left*, Internet-hosted until 2010) and actual photograph taken from the same viewing position. *Courtesy* Michael Gruber, Microsoft

showing a New York building assembly in a visualization that is being compared to a real photograph of the scene taken from a ship.

The developments began about 40 years ago. Phase I took place in the 1970s with the massive digitization of national and commercial paper maps with the ability to copy, transfer and sell maps at low variable cost in an early form of e-commerce. One was vectorizing the raster images coming off very accurate cartographic scanners, each map color separate resulting in arcs and nodes, small data quantities and in automated changes of scale. Contour lines were vectorized as well and converted into digital elevation models DEMs. The map separates and elevation information made it possible that the 1960-ideas about a Geographic Information System GIS (Tomlinson 1967) actually got implemented.

Phase II was inspired by the idea of road maps for trip planning which needed national road maps augmented by addresses and information about one-way streets and turn information, ETAK was an early 1983-innovator, later merged into Teleatlas. This company, together with Navteq, evolved into the dominating commercial Geo-data providers.

Phase III may be associated with the advent of the Internet as the enabler of widespread commerce with Geodata. In lieu of shrink-wrapped Geodata packages supporting mapping applications, the data now were conveniently presented online. Another application emerged for trip planning using Geodata via the Internet. The early innovator was MapQuest.

But it took Phase IV with instant satellite geo-positioning to not only plan travel, but to navigate vehicles in real time, however using Geo-data stored on board a car, thus "Internet-free". The subsequent introduction of maps in Internet search engines may be considered a separate Phase V. This started the idea of a Location Aware Internet since 2005. We can visit any place on the globe on our computer, wherever we are. Any location-specific information associated with a search can be presented on a map on the computer monitor. Google, Microsoft, Ask, Yahoo all embarked on such services, with Google-Maps and Google-Earth having the greatest penetration.

Phase VI is now emerging as mobile communications transit into the ubiquitous smart phone. Not vehicles are navigating, but people (things?) are. The

Internet-of-Things, Ambient Living and location-based applications of social networks represent vast opportunities from knowing at all times where persons and things are. Nokia was an early adopter, rapidly followed now by Google and other Internet search engine providers. Apple-Maps has been the most recent entry into this service. With Facebook and Google competing to own WAZE, we see Facebook's intentions to also be a player.

Associated with each evolutionary Phase was and is an emergence of new research trends and scientific-technical conferences, as well as new businesses.

The 40-year-transition from paper maps to today's "Internet maps" (Peterson 1997, 2008) resulted in augmented street-maps used for car navigation, adding to these the terrain shape in the form of the Bald Earth, augmenting this information by photographic texture from orthophotos, and supporting the system by a variety of photographic data from the air, specifically from vertical and oblique looking cameras. Large areas of the industrialized World are fully presented on the Internet when calling up the websites maps.google.com or www.bing.com/maps, and a number of regional Internet mapping services, for example the French www.geoportail.fr or the German www.klicktel.de/kartensuche.

Mapping data on the Internet had significant consequences in the form of the GIS, the role of amateurs or "neo-geographers" (Goodchild 2008), easy mix of 2D and 3D data, augmentation of geometrically accurate mapping data by casual amateur photography, very large format digital imagery and very small cameras, use of unmanned vehicles, automated processing using many more than just two images per terrain point, and visualization of 3D, sometimes even 4D Geo-data using the tools of computer graphics and mixed reality.

The significance of location for new markets has increased the budgets for research and innovation in mapping related fields. Computer vision, 3D urban modeling, augmented reality have evolved into significant sources of innovation. Early city models were of interest in the context of urban warfare. In Europe, this has been a research topic since the mid 1990s (Dang et al. 1993; Grün et al. 1995; Förstner and Weidner 1995; Gruber et al. 1995). Early implementations of 3D within Internet-supported search was at Bing/Maps (then Virtual Earth) and had buildings represented as triangulated point clouds. Photo texture served to add visual detail and embellishment. Increasingly, objects in such 3D models get interpreted, for example as trees, circulation spaces, buildings. The Bald Earth gets used as a geometric basis onto which on places the man-made 3D objects (Leberl and Gruber 2009b; Kluckner and Bischof 2010).

Current research seeks to interpret the objects of a 3D model. The initial 3D-location awareness represented "eye candy", and is not the basis of the search itself. The 2D content of a street map contains address codes and can be searched. The 3D model should be searchable but is not at this time. Interesting questions would be the number of windows or floors, the orientation with respect to the sun, the built-up surface area, the extent of impervious terrain, the type of roof. Extracting such information from existing imagery and data bases is a challenging research topic (Leberl et al. 2009, 2010b).

Opportunities are emerging from the Internet-of-Things and Ambient Intelligence with their need for location awareness (O'Reilly and Batelle 2009). It was already in 1991 that Marc Weiser authored his much quoted prediction for computing in the twenty-first century and postulated that *location* will be one of two issues of crucial importance: "*ubiquitous computers must know where they are*" (Weiser 1991).

Object tracking is being accomplished by sensors/RFIDs and by embedded computing so that a 3D location is available with an Internet UID for each object. For this location to make sense, one will need a model of that World at a detail commensurate with the things surrounding us humans. One often speaks about "human scale detail" and implies detail in the decimeter to sub-decimeter (thus centimeter) range.

2 The Internet Since the 1980s

"*How does the Internet work*"? It all started as an application of then-existing telecommunication systems. The grandfather was the 1969-introduction of the ARPANET in the US. With manual help, one computer makes a telephone call to another computer and upon establishment of the connection data get sent over the telephone line. The older among us will remember the typical dial-up sound in preparation of the acoustic coupling (Fig. 2).

This basic link between two computers is the essence of the Internet, although today one refers to the Internet as a *network of computer networks*, exceeding today 1 billion connected computer servers. While the word itself is from the 1880s, its first modern use was documented in 1974. Great strides have been made in many aspects of these basic concepts, be it speed, data volume, fiber optic connections, mobile systems. Essentially, the Internet is a beneficiary of telecom innovations. Just as a medieval dirt road connecting two cities is the basic idea of traffic lines, so are the advances to freeways and train lines the analogy to the advances in telecommunications. From the early 56 kbits per second, one finds today data rates in excess of 100 Gigabits per second.

Fig. 2 Acoustic coupling of two computers via a telephone line—www.youtube.com/watch?v=ychSsyn4xPs

Fig. 3 Fiber bundle (*left*) and aso-called "core router" (*right*), example of a current contemporary CISCO Internet-device (from en.wikipedia.org/wiki/Core_router)

One explanation of the Internet may be helped by a look at the industry providing the infrastructure. There are providers of fiber optic lines, computer devices, security devices, etc. A leading industry player is CISCO, founded in 1984 by two staff members of Stanford University (CISCO is short for San Francisco) and grew to a market capitalization of over USD 500 billion by 2000, the most valuable company in the World. CISCO is the ultimate Internet-infrastructure provider (see Fig. 3).

Internet-software deals with both the infrastructure and the vast applications. Entire new industries have emerged, beginning with software as a product, communications software, web portals, search engines, social media, etc. Corporate businesses have been and are being created to this day from zero to multi-billions in a span of a very few years. The most visible corporate values created recently are by Google and Facebook. A very recent success story is WAZE, started in 2008 and sold in 2013 for a reported USD 1.3 billion.

40 % of humanity uses the Internet today. With the takeover of mobile communications by the smart phone, the Internet has become part of cell-based telecommunications and therefore is expected to be in use by essentially all of humanity by the year 2020.

3 Digital Decades

Given that the history of digital computing goes back to the inventions by Konrad Zuse in 1941, it surprises that the concept of a "*First Digital Decade*" was promoted by Bill Gates in January 2008, ~ 60 years later. This 1st digital decade is characterized by (1) the total install base of personal computers in excess of 1 billion, (2) availability of cellular telephony to more than 40 % of humanity, (3) growth of broadband services from 0 to 250 million users and (4) the transition from film to digital cameras, and therefore from film to the power of software. Under this definition, we currently live in the second digital decade. The cell phone penetration has reached 6 billion by the Fall of 2012, the smart phone is expected to completely have replaced traditional cell telephony by 2020, computing is in a transition to becoming wearable, with each person carrying multiple computing devices on his or her body. 30 % of humanity, thus 2.1 billion people, today accesses mobile broadband, 10 times more than did in 2008.

The dynamics of computing has been reflected by very disruptive innovations such as Internet search or social media. It also has experienced gradual evolution, such as a 12.5 million improvement of the cost-performance of digital storage over a period of 30 years from 1975 to 2005.

All this is being driven by the paradigm implied by Moore's Law with its price performance improvement by a factor of 2 each 1.5 years. This produces an improvement by a factor 100 across a 10-year period and by 1 million across 30 years.

Growth of all things related to computing promises to continue. We currently see the vast implementation of the RFID (Radio Frequency Identification Devices) in industry, with an Internet address for each "thing" evolving into the Internet-of-Things IoT. Not only human users, businesses and organizations have Internet addresses, but all things, animals, even vegetation, are expected to be found with their separate addresses on the Internet.

4 From Digital Maps to the Location-Aware Internet

We have already presented a 6-phase-evolution from paper maps to today's Internet- and GPS-based ubiquitous location awareness. The field of mapping was an early user of fairly massive computing resources by converting traditional paper maps and map separates into raster images. Some of the early innovations in the 1970s were driven by the need of weapons systems such as the Cruise missile.

Search on the Internet became location-aware by associating Internet-based maps with search results, and integrating directions and navigation with the Internet. In 2005, Google introduced maps.google.com, Microsoft maps.live.com which later was relabeled as www.bing.com/maps. Massive Geo-data need to be provided to large numbers of users. The data structure is based on triangles or on square tiles (Fig. 4), and an intelligent navigation through those tiles at precomputed levels-of-detail. The augmentation of these mapping systems by 3D urban building models started in Nov-2006 with Microsoft's making available Virtual Earth in 3D (Paul 2006). 3D is being advertised by Apple and Google, maps also get augmented by imagery taken systematically from the air by vertically and oblique-looking cameras, as well as from the street level, or haphazardly by amateurs in the form of Community Photo Collections (Gösele et al. 2010). Location on the Internet is increasing in relevance as smart phones offer pedestrian location applications. This cannot be well separated from the broad field of Location Based Services (Gartner and Rehrl 2008).

While applications like Bing-Maps or Google Earth are currently driven to pull the public into using the associated Internet-search system via attractive location awareness, there is a deeper justification in light of the emerging opportunities created by the Internet-of-Things and Ambient Intelligence. Weiser's sketch of the future has morphed from ubiquitous computing to ambient intelligence. We expect the geometry of the elementary parts of an entire urban environment to be
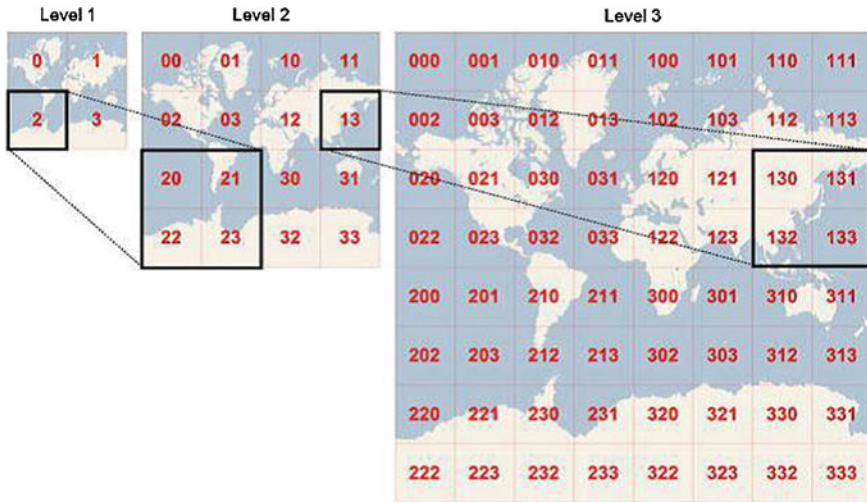
Fig. 4 A tiling scheme for fast access to geographic data (from Kröpfl 2013)

available at a human scale to be searched, we foresee that this becomes the basis for locations of fixed sensors, or moving GPS-tracked sensors, to read the RFID- and other tags of goods to place these inside buildings, even inside individual rooms. In analogy, triangulated cell phones as well as tags can also place persons inside buildings and rooms. A semantically interpreted 3D city model could help find one's briefcase, an errant person, the nearest copying machine and will affect the use and behavior of all sorts of computer-driven gadgets.

Considerable infrastructure investments are being made to create and maintain large Geo databases. As an example, Fig. 5 shows how a container-based data processing center in support of the location-aware Internet system of Microsoft gets set up. The 24,000 CPU cores can process 200,000 aerial photographs in 1 day, covering an ortho-product of 1,000,000 km^2 and resulting in 250 TB of new data. This equals a throughput of 37 DVDs per-minute or 51,000 DVDs total.

5 Towards Semantically Interpreted Urban Models

"*Eye-candy*" is a way to denote the weakness of current imagery and 3D models on the Internet in use to augment search results. Such data cannot be searched. We cannot ask the Internet about the number of floors in a 3D building model, for example. The initial purpose of maps in searching the Internet simply was to attracted "eyes" to a specific search engine, and to keep the user away from competing engines. However, research into the creation of interpreted 3D models is happening and addressing the conversion of imagery and 3D urban models into a
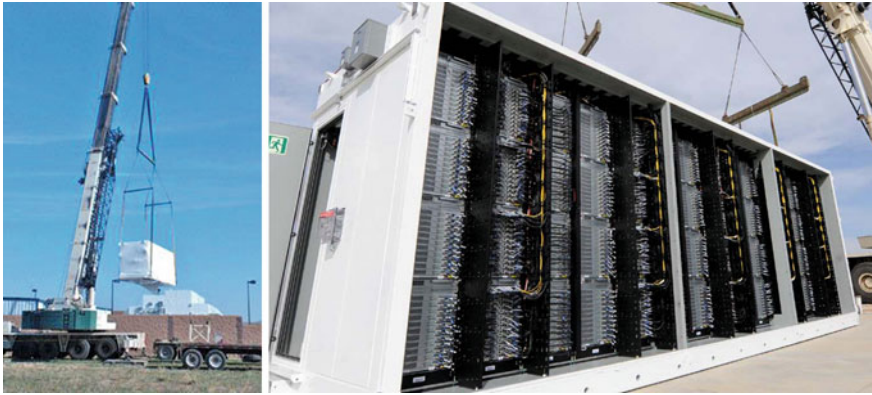
Fig. 5 Setting up a data processing center with multiple containers in Boulder, Colorado, to support the BING/Maps initiative. "Lights-out" compute and storage management with 24,000 CPU cores, 71 PB Storage, 52,504 HDDs (from Walcher et al. 2012)
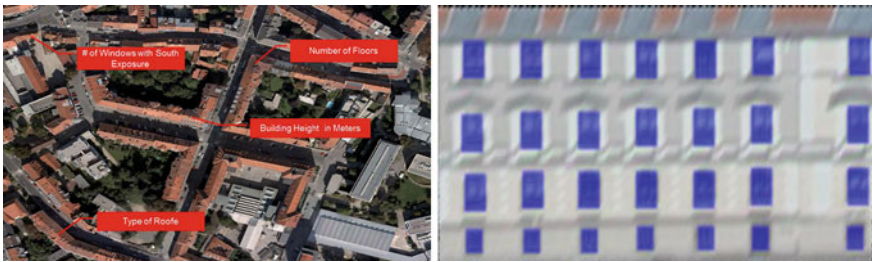
Fig. 6 Searchable interpreted 3D model of buildings in Graz (*left*) and a detail of an interpreted façade with its windows (from Meixner and Leberl 2011a, b)

data base of meaningful objects (Leberl et al. 2009, 2010b). Figure 6 illustrates the basic concept of an interpreted urban model and typical questions one might ask of such a model.

6 The Role of Community Photo Collections

The concept of a "map" is evolving. From the millennia-old paper map presenting an image of the World using symbols, scale and a well-defined geometric accuracy, we now see data systems on the Internet without much respect for scale or geometric accuracy, and with a plethora of data types far exceeding the simple vector-type line map. The most dramatic augmentation of the map, when presented on the Internet, is the collection of everybody's photographs attached to map locations. We can therefore see locations as we have in the past on maps, but we

Fig. 7 A systematically created urban panorama is the backdrop over which a user-contributed photo gets automatically draped (from Kröpfl et al. 2012)

also can see how the location appears from the street level, and through the cameras of any number of people willing to share their photographic records visa image data bases such as FLICKR, Photobucket, Panoramio.

The use of consumer-contributed imagery is being illustrated in Fig. 7. A BING/MAPS location using the streetside image panorama gets augmented by user-contributed historical photography of an area in Seattle. Matching is being accomplished automatically (Kröpfl et al. 2012).

7 Golbal Ortho

For a global Internet search system, a global map coverage is required. The vector-type infrastructure data typically have been collected over time by commercial operations, most visibly by Navteq and Teleatlas, and they sell them to various users, be they providers of car navigation or Internet search services. Recent alternative approaches assemble fresh road data, for example from the use of GPS tracks collected via cars or smart phones. Image-maps are not globally available, except for satellite imagery. National orthophoto coverage often exists and can be licensed from National mapping agencies. Those data sets may be outdated, license fees might be high, the quality and geometric detail might be compromised.

Both Google and Microsoft have therefore embarked on vigorous efforts to develop their own high resolution orthophoto coverage in all World regions where this is not being denied via national sovereignty restrictions. Microsoft's project is denoted as the Global Ortho (Walcher et al. 2012). A special aerial camera was designed for most efficient aerial data collection and is denoted as UltraCam-G (Fig. 8, Leberl et al. 2012). The initial coverage was of the entire continental USA and industrial Western Europe, and an area in excess of 10 million km^2 has been produced over a period of less than 3 years. The orthophotos have a geometric
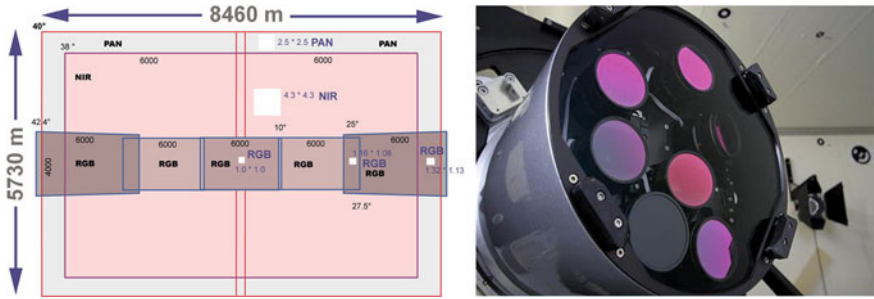
Fig. 8 The UltraCam-G for use in the Global Ortho project. Eight optical heads collect two panchromatic image tiles, one infrared and five color image tiles (Leberl et al. 2012)

Fig. 9 Global orthophoto at a pixel size of 30 cm (*left*) and DEM with postings at 1 m interval (*right*). (Walcher et al. 2012)

resolution of 30 cm, are produced in red–green–blue and near infrared. Associated with the orthophotos are digital elevation models (see Fig. 9). The global orthophoto is used by the Internet search engine BING. The data can also be purchased via the Geo-data services from Digital Globe.

8 Three Dimensions

The human is immersed in a 3D life environment. The Internet therefore also will want to mimic that immersion, as locations increasingly support Augmented Reality approaches to location-experiences.

Clearly, 3D urban models still are in their infancy, and most current systems are more experiment than application. 3D may have various connotations. One is the actual collection of 3D source data such as point clouds from LiDAR or multi-image photogrammetry, the subsequent development of building models from those point clouds. Another is the 3D experience via streetside imagery of building façades and associated vegetation that present the environment's 3D features in a series of 2D views.
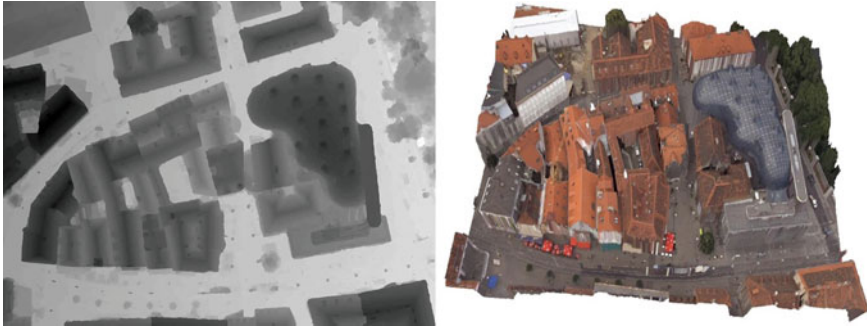
Fig. 10 Automatically produced 3D point cloud from 15-times overlap aerial photography (*left*) and photo textured (*right*). Downtown Graz

At issue are various technological evolutions. First is of course the ability to automate the creation of dense and geometrically accurate point clouds from highly overlapping aerial and streetside photo-graphy. This technology has come a long way and is fairly well developed (Frahm et al. 2010, also Fig. 10). High performance GPU-based computing makes it feasible to use fairly complex algorithms for fully automated point cloud computations. This combines with previously unthinkable image quality at high overlaps and no variable costs for digital imaging. The point clouds may be at densities of perhaps 50,000 points per m^2. These need then to get interpreted into the individual 3D objects of each façade in the form of windows, doors, eaves, columns, decorations, etc. Second is the clarification of relative merits of dense photogrammetric versus LiDAR point clouds This question is not well researched (Leberl et al. 2010a). Third is the ability of matching data and extracted observations from the air, from systematically collected streetside imagery and from accidental imagery found in Community Photo Collections. This type of matching presents great challenges (Kröpfl 2013, Kröpfl et al. 2012). Geometric resolutions vary greatly, poses may be known very inaccurately, aspect angles may be fairly random, illumination and radiometry are undetermined.

All Internet mapping sites offer some form of 3D. Microsoft initially made it a mainstay of its system in 2005–2006, but abandoned the initiative in 2010 for a lack of user response. Google relies on user content to add 3D data to Google Maps. Apple advertises 3D models in its maps.Apple.com site, and this places 3D center stage.

While Internet-search may be the most visible and also the initial "killer-application", there are others: city planning, virtual tourism, disaster preparedness, military or police training and decision making (Willkomm 2009), or 3D car navigation (Strassenburg-Kleciak 2007). The German GIS-world takes the 3D urban models in such high regard that an all-encompassing CityGML data standard has been developed, going from a coarse level 0 to a level 4 including building interiors. The international OGC Open Geospatial Consortium has adopted this standard (Kolbe et al. 2009).

9 Some Current Research Themes

The interpretation of images of the human habitat, be they aerial, streetside, indoors, systematic or accidental, remains a major topic of research. Whether this addresses the original 2D images or 3D point clouds with the images, the goal always is a description of the scene content in 3D. Such interpretations are a particular challenge when the imagery is "accidental", thus extracted from a CPC or a new photo taken by an Internet user. (Meixner and Leberl 2011a, b) has focused on the interpretation of 3D buildings to extract the number of floors and widows, the type of roof, the orientation of a building vis-à-vis the sun. Kluckner and Bischof (2010) interpreted the objects of an urban scene such as roads, buildings, vegetation types. Kröpfl et al. (2012) is interested in streetside imagery, especially accidental images taken by amateurs, and matching those with pre-existing 3D models or urban panoramic images.

Of concern is privacy when considering centimeter-type image resolution. This addresses faces of people and car license plates. Streetside images in Internet systems typically get published after such areas have been identified and made anonymous (Fig. 11).

Sensor platforms used to be fairly well defined in the form of aircraft flying aerial cameras. Today, the air may carry Unmanned Aerial Vehicles, and if they are very small, so-called MAVs. Wendel et al. (2012) has been successful in using such platforms for urban 3D mapping using large numbers of irregularly arranged aerial photos with irregular poses (Fig. 12).

Such MAV-based images do provide high resolution photos of building details, but multiple internally connected image blocks may not connect across larger

Fig. 11 Privacy detection by marking areas assumed to contain private information, and subsequently making these areas unintelligible (Kröpfl et al. 2012)

Fig. 12 MAV-based imaging into an irregular assembly of aerial images, and extraction of a dense point cloud of the clock tower in Graz (from http://www.youtube.com/watch?v=vYXhVG7VMy0)

distances. Source images may produce sepa-rate point clouds in need of fusion. This "fusion" represents an important topic of research (Fig. 13).

Oblique aerial photography has grown for measurements about real estate property by non-expert users. Pioneered by Pictometry, it has found numerous followers. Most recently an aerial camera for both vertical as well as oblique aerial photography has been introduced by Microsoft under the name UltraCam Osprey (www.microsoft.com/ultracam/en-us/UltraCamOsprey).
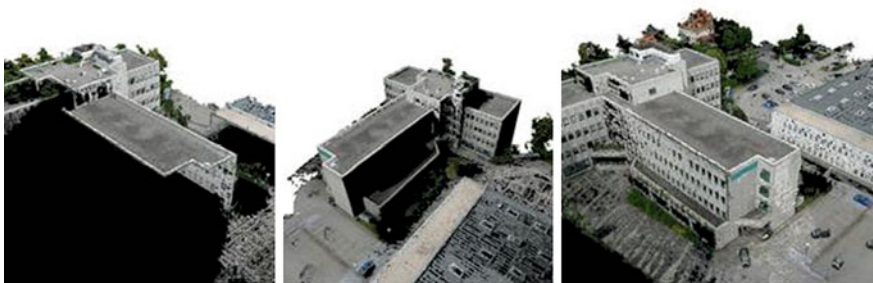
Fig. 13 Point clouds from two separate sets of MAV-based image blocks (*left* and *middle*) of one terrain object. These point clouds get fused into a single point cloud to the *right* (from Wendel et al. 2012)

10 An Outlook: A Virtual Centimeter World Model

One wishes for an urban model to read all shop signs, to move around a building's interiors and possibly around merchandise, to inspect suspended wires and details of the urban streetscape such as the elevation differences of sidewalks and building entries. This calls for detail in the range of centimeters. And all this is to be provided by fully automated processes. Figure 14 illustrates recent 3D Internet-models as examples of current capabilities.

Graz is a medium size city with 250,000 people, 25,000 buildings and 1,000 km streets. If every building were photographed by owners on ∼40 photos, we would have 1 million images with a "few centimeter" pixels. If in addition we had some MAVs fly along the street canyons to cover the road surface, roofs, courtyards and public spaces, at an interval of a few meters, we would have an added assembly of perhaps 250,000 aerial photos. Some systematically collected streetside images and LiDAR data from a car would collect another set of 10 photos every 5 m or so. This will add another ∼2 million photos. If each of those photos were to cover an array of 4 K by 3 K pixels, we would have collected a total of 39 TB for such a city. That is the assembly of source data, not yet the resulting model data.

Processing a million of such photos can be achieved in a day, as demonstrated by Frahm et al. (2010), Snavely et al. (2008) and Gösele et al. (2010). The 3D model, augmented by photo texture and interpreted façade objects, will represent a cm-type model. This is feasible today with today's sensors, computing resources and storage systems without excessive cost. While this can be done today, it has yet to be demonstrated. We are waiting for the first mayor of any city to issue the request for each building owner to photograph it and to upload the result.

4,000 cities like Graz would be home to a billion people. Half the World's population lives in cities, half in rural areas and smaller towns. Each city would be subject to its own data collection and processing. 4,000 cities like Graz will result in a 156 Exabyte source data set. The quantity is driven by the source data of high overlap photos, covering each terrain and building façade point ∼30 times or so.
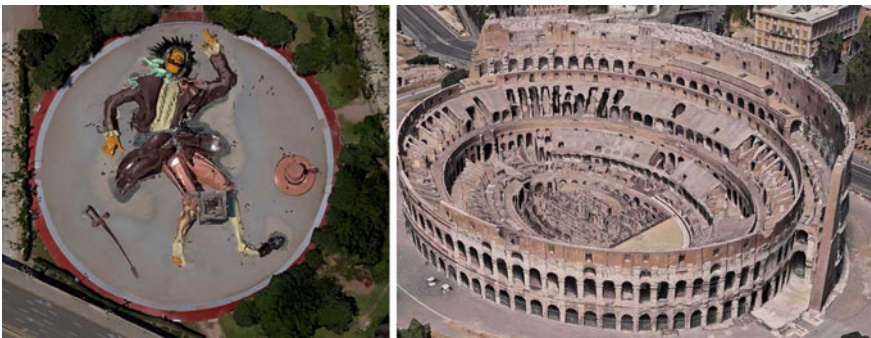
Fig. 14 *Left*—A children's play area from a 3D model of Valencia, Spain, *courtesy* Microsoft Photogrammetry. *Right*—Rome coliseum, from Apple maps

The resulting 3D models would be very much more compact and consist of single clouds of points per building, and single photo textures per building façade. The multi-image redundancy would remain in the source data only.

Given that we today have the technology in hand to accomplish such centimeter-type model of urban spaces, we may ask the question *how long it will take for people to catch on* and to decide to actually create such data sets. This requires political will and a champion to get the ball rolling. It does not—and this is important to realize—require any breakthroughs in technology.

References

Dang T, Jamet O, Maître F (1993) Interprétation et Restitution Automatique de Bâtiments en Milieu Péri-Urbain. Revue Française de Photogrammétrie et Télédétection 131(3):3–12

Förstner W, Weidner U (1995) Towards automatic building reconstruction from high resolution digital elevation models. ISPRS J photogramm Remote Sens 50(4):38–49

Frahm J-M, Georgel P, Gallup D, Johnson T, Raguram R, Wu C, Jen Y, Dunn E, Clipp B, Lazebnik S, Pollefeys M (2010) Building rome on a cloudless day. European conference on computer vision ECCV 2010, Crete, Greece

Gartner G, Rehrl K (eds) (2008) Location based services and telecartography II. From sensor fusion to context models. 5th international conference on location based services and telecartography, 2008, Salzburg (Lecture notes in geoinformation and cartography, 2009, XXIX, p 456. ISBN: 978-3-540-87392-1)

Goodchild M (2008) Assertion and authority: the science of user-generated geographic content. Proceedings of the colloquium for Andrew U. Frank's 60th birthday. GeoInfo 39. Department of Geoinformation and Cartography, Vienna University of Technology

Gösele M, Ackermann J, Fuhrmann S, Klowsky R, Langguth F, Mücke P, Ritz M (2010) Scene reconstruction from community photo collections. IEEE Comput 43(6):48–53

Gruber M, Pasko M, Leberl F (1995) Geometric versus texture detail in 3D models of real world buildings. In: Grün A, Kübler O, Agouris P (eds) Man-made structures from aerial and space imagery. Birkhäuser, Basel, pp 189–198

Grün A, Kübler O, Agouris P (eds) (1995) Man-made structures from aerial and space imagery. In: Proceedings of an international workshop at Monte Verità, Ascona, Switzerland. Birkhäuser-Verlag, Basel, p 250

Kluckner S, Bischof H (2010) Large-scale aerial image interpretation using a redundant semantic classification. ISPRS Archives, vol XXXVIII, Part 3A, pp 233–238

Kolbe T, Nagel C, Stadler A (2009) CityGML-OGC standard for photogrammetry? In: Proceedings of the photogrammetric week'09. Wichmann-Heidelberg Publishers, Heidelberg, pp 265–277. ISBN 978-3-879-7-483-9

Kröpfl M (2013) The role of imagery for web-based mapping applications. Dissertation, Graz University of Technology

Kröpfl M, Buchmüller D, Leberl F (2012) Online maps and cloud-supported location-based services across a manifold of devices. ISPRS Ann Photogram Remote Sens Spat Inf Sci I-4:151–156

Leberl F (2007) Die automatische Photogrammtrie für Microsoft Virtual Earth Internationale Geodätische Woche Obergurgl. Chesi/Weinold(Hrsg.), Wichmann-Heidelberg-Publishers, Heidelberg, pp 200–208

Leberl F, Gruber M (2009a) Ortsbewusstsein im Internet—von 2-dimensionalen Navigationshilfen zur 3-dimensionalen mixed reality. Tagungsband der 15. Geod. Woche Obergurgl, Wichmann-Verlag, pp 67–69. ISBN 978-3-87907-485-3

Leberl F, Gruber M (2009b) 3D-models of the human habitat for the Internet. In: Proceedings of VISIGRAPP-2009, Lisbon, Portugal, INSTCC-Portugal, vol IS, pp 7–15. ISBN 978-989-8111-69-2

Leberl F, Kluckner S, Bischof H (2009) Collection, processing and augmentation of VR cities. In: Proceedings of the photogrammetric week 2009, Wichmann, Stuttgart, pp 251–264. ISBN 978-3-879-7-483-9

Leberl F, Irschara A, Pock T, Meixner P, Gruber M, Scholz S, Wiechert A (2010a) Point clouds: LiDAR versus 3D vision. Photogram Eng Remote Sens 76:1123–1134

Leberl F, Bischof H, Pock T, Irschara A, Kluckner S (2010) Aerial computer vision for a 3D virtual habitat. IEEE Comput 43(6):24–31

Leberl F, Gruber M, Ponticelli M, Wiechert A (2012) The ultracam story. Intl Arch Photogram Remote Sens Spat Inf Sci XXXIX-B1:39–44

Meixner P, Leberl F (2011a) 3-dimensional building details from aerial photography for Internet maps. Remote Sens 3:721–751

Meixner P, Leberl F (2011b) 3-dimensional building details from aerial photography for Internet maps. Remote Sens 3:721–751

O'Reilly T, Batelle J (2009) Web Squared: Web 2.0 five years on. O'Reilly Media Inc. www.web2summit.com

Paul R (2006) Microsoft launches Virtual Earth 3D to try and take on Google Earth. http://www.earthtimes.org/articles/show/10224.html (posted 7 Nov 2006)

Peterson M (1997) Trends in Internet map use. In: Proceedings of the 18th international cartographic conference, Stockholm, Sweden, vol 3, pp 1635–1642

Peterson M (2008). Trends in Internet and ubiquitous cartography. Cartograph Perspect 61:36–49

Snavely N, Seitz SM, Szeliski R (2008) Modeling the world from Internet photo collections. Int J Comput Vis 80(2):189–210

Strassenburg-Kleciak M (2007) Photogrammetry and 3D Car navigation. In: Fritsch D (ed) Proceedings of 51st photogrammetric week. Wichmann, Berlin, pp 309–314

Tomlinson RF (1967) An introduction to the geo-information system of the Canada Land Inventory. ARDA, Canada Land Inventory, Department of Forestry and Rural Development, Ottawa

Walcher W, Leberl F, Gruber M (2012) The Microsoft global ortho program. ISPRS Ann Photogram Remote Sens Spat Inf Sci I-4:53–58

Weiser M (1991) The computer for the 21st century. Scientific American, special issue on communications, computers, and networks, vol 265, No. 3, September

Wendel A, Hoppe C, Bischof H, Leberl F (2012) Automatic fusion of partial reconstructions. ISPRS Ann Photogram Remote Sens Spat Inf Sci I-3:81–86

Willkomm P (2009) 3D GDI—Automationsgestützte Erzeugung und Verteilung landesweiter Gebäudemodelle aus Laserdaten. In: Proceedings of 14th Münchner Fortbildungsseminar GIS, Technische Universität München, unpaginated DVD

Part V
Planetary Mapping Issues

Jacobi Conformal Projection of the Triaxial Ellipsoid: New Projection for Mapping of Small Celestial Bodies

Maxim V. Nyrtsov, Maria E. Fleis, Michael M. Borisov and Philip J. Stooke

Abstract In this paper a new technique for recalculating geographic coordinates of a triaxial ellipsoid to elliptical and then to rectangular coordinates of the Jacobi conformal projection is considered. Coordinate lines of the elliptical system and the cartographical grid with the parallels passing through the circular points on the Jacobi projection are shown. This new technique allows us to achieve the conformal mapping of small celestial bodies. A map of asteroid 25143 Itokawa in the Jacobi conformal projection, the first ever published, and a map of asteroid 433 Eros created by the authors in the transverse conformal cylindrical projection of a triaxial ellipsoid are presented for comparison. Asteroids 25143 Itokawa and 433 Eros are near-Earth objects.

1 Introduction

Carl Jacobi in his lectures (1842–1843) at the University of Königsberg proposed the conformal projection of the triaxial ellipsoid to the plane. The lectures were written down by C.W. Borchardt and published by A. Clebsch in 1866 (Jacobi 1866). They also were translated into many languages particularly in English (Jacobi's lectures 2009). We used a Russian translation made by O.A. Polosukhina

M. V. Nyrtsov (✉)
Moscow State University of Geodesy and Cartography (MIIGAiK), Moscow, Russia
e-mail: nyrtsovmaxim@gmail.com

M. E. Fleis · M. M. Borisov
Institute of Geography, Russian Academy of Sciences, Moscow, Russia

P. J. Stooke
The University of Western Ontario, London, ON, Canada

in 1936 and edited by N.S. Koshlyakov, a specialist in the field of partial differential equations (Jakobi 1936). The interpreters tried to translate the text not only as accurately as possible, but also to keep as much as possible the original features of the language and the character of the descriptions.

This is important because Jacobi's lectures are interesting not only for their results, but also for their depiction of the author's thoughts. For example, we are interested in a map projection obtained as a result of applying the elliptic coordinates to the derivation of the shortest line equation on the triaxial ellipsoid. Many of the materials presented in the lectures were used by mathematicians in development of the theory of surfaces. In the 1970s and 1980s interest in map projections for triaxial ellipsoids began to grow among cartographers in connection with the problem of mapping of small celestial bodies.

The Jacobi conformal projection is considered in Bugaevskiy (1999), but its formulae were found to be inconvenient for practical use because of difficulties in calculating of the integrals. In this paper we consider the technique of recalculating geographic coordinates of a triaxial ellipsoid to elliptical and then to rectangular coordinates of the Jacobi projection and mapping in this projection.

2 Derivation of Jacobi Projection

In his 26th lecture Jacobi introduced the elliptic coordinates $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n$ for the multidimensional case using the equation

$$\frac{x_1^2}{a_1 + \lambda} + \frac{x_2^2}{a_2 + \lambda} + \cdots + \frac{x_n^2}{a_n + \lambda} = 1. \tag{1}$$

Later they became known as Lamé coordinates. He examines them analytically. For our purpose the important point of this lecture is the relationship between x and y values and the relationship between the squares of their differentials. In the 27th lecture the geometric interpretation of the results of the previous lecture, applied to the plane and three-dimensional space is given. For the three-dimensional case different ranges of λ changes correspond to three systems of confocal quadric surfaces (hyperboloid of one sheet, hyperboloid of two sheets, triaxial ellipsoid). In the three-dimensional space one hyperboloid of one sheet, one hyperboloid of two sheets and one triaxial ellipsoid are going through each point. These surfaces intersect each other at a right angle. According to this, the square of the element of arc of arbitrary curve, expressed in terms of the differentials of elliptic coordinates, doesn't contain the multiplication of differentials of two different values. Geometric interpretation of elliptic coordinates is considered in more detail in Kagan (1947).

The 28th lecture considers an orthogonal system of elliptic coordinates on the surface of the triaxial ellipsoid. Equation

$$\frac{x_1^2}{a_1 + \lambda} + \frac{x_2^2}{a_2 + \lambda} + \frac{x_3^2}{a_3 + \lambda} = 1, \tag{2}$$

where $a_1 < a_2 < a_3$ for given values of rectangular coordinates has three real roots: $\lambda_1 > \lambda_2 > \lambda_3$. Root λ_1 corresponds to the ellipsoid, root λ_2 corresponds to a hyperboloid of one sheet, root λ_3 corresponds to a hyperboloid of two sheets.

For the surface of a given ellipsoid the value of λ_1 is constant. After obtaining the formula for the square of the element of arc of an arbitrary curve on the ellipsoid

$$ds^2 = \frac{1}{4} \frac{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)}{(a_1 + \lambda_2)(a_2 + \lambda_2)(a_3 + \lambda_2)} d\lambda_2^2 + \frac{1}{4} \frac{(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)}{(a_1 + \lambda_3)(a_2 + \lambda_3)(a_3 + \lambda_3)} d\lambda_3^2.$$

Jacobi determines the ratio between the element of arc on the plane (in projection) $d\sigma$ and on the ellipsoid $d\sigma = \frac{2}{\sqrt{\lambda_2 - \lambda_3}} ds$. Also Jacobi determines formulae of the conformal projection

$$u = \int d\lambda_2 \sqrt{\frac{\lambda_2 - \lambda_1}{(a_1 + \lambda_2)(a_2 + \lambda_2)(a_3 + \lambda_2)}}$$

$$v = \int d\lambda_3 \sqrt{\frac{\lambda_1 - \lambda_3}{(a_1 + \lambda_3)(a_2 + \lambda_3)(a_3 + \lambda_3)}}. \tag{3}$$

Here u and v are rectangular coordinates on the plane.

Note. Values $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n$ are measured in units of x^2, value ds is measured in units of x (units of the triaxial ellipsoid axes), $\frac{2}{\sqrt{\lambda_2 - \lambda_3}}$ is measured in units of $\frac{1}{x}$, values $d\sigma$, u, v are dimensionless.

3 Coordinate Systems and Coordinate Lines

Later rectangular coordinates on the plane of projection we will sign as x_{proj} and y_{proj}. We have also changed the signing of elliptic coordinates, because the value λ is usually used for longitude. Because a traditionally cartographic grid is used for mapping, we take angular planetocentric coordinates (ϕ—latitude, λ—longitude) as the initial (see Fig. 1).

In a new (more usual) values Eq. (2) becomes

$$\frac{x^2}{a^2 - u} + \frac{y^2}{b^2 - u} + \frac{z^2}{c^2 - u} = 1, \tag{4}$$

where
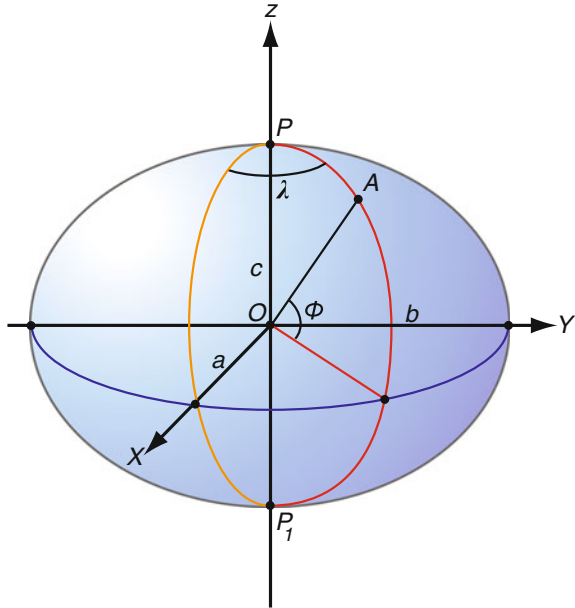
$c^2 = a_1 + \lambda_1$, $\quad b^2 = a_2 + \lambda_1$, $\quad a^2 = a_3 + \lambda_1$—squares of semi-axes of the triaxial ellipsoid, $x = x_3$, $\quad y = x_2$, $\quad z = x_1$—three-dimensional rectangular coordinates, $u = \lambda_1 - \lambda$.

For given three-dimensional rectangular coordinates elliptic coordinates are getting from Eq. (4) relatively to u

$$x^2 (b^2 - u)(c^2 - u) + y^2 (a^2 - u)(c^2 - u) + z^2 (a^2 - u)(b^2 - u)$$
$$= (a^2 - u)(b^2 - u)(c^2 - u)$$

$$u^3 + u^2 (x^2 + y^2 + z^2 - a^2 - b^2 - c^2)$$
$$+ u(-x^2 b^2 - x^2 c^2 - y^2 a^2 - y^2 c^2 - z^2 a^2 - z^2 b^2 + a^2 b^2 + b^2 c^2 + a^2 c^2)$$
$$+ x^2 b^2 c^2 + y^2 a^2 c^2 + z^2 a^2 b^2 - a^2 b^2 c^2 = 0.$$

On the surface of the triaxial ellipsoid
$x^2 b^2 c^2 + y^2 a^2 c^2 + z^2 a^2 b^2 - a^2 b^2 c^2 = 0$, that is one of the roots $u_3 = 0$.
Solve a quadratic equation with non-zero u

$$u^2 + u(x^2 + y^2 + z^2 - a^2 - b^2 - c^2)$$
$$- x^2 b^2 - x^2 c^2 - y^2 a^2 - y^2 c^2 - z^2 a^2 - z^2 b^2 + a^2 b^2 + b^2 c^2 + a^2 c^2 = 0$$

and get two roots

$$u_1 = \frac{-p + \sqrt{p^2 - 4q}}{2} \text{ and } u_2 = \frac{-p - \sqrt{p^2 - 4q}}{2}, \text{ where}$$

$$p = x^2 + y^2 + z^2 - a^2 - b^2 - c^2$$
$$q = -x^2 b^2 - x^2 c^2 - y^2 a^2 - y^2 c^2 - z^2 a^2 - z^2 b^2 + a^2 b^2 + b^2 c^2 + a^2 c^2.$$

Sign $u_1 = u$, $u_2 = v$. Integrals (3) become

$$x_{proj} = \int_{b^2}^{u_i} \sqrt{\frac{u}{(c^2 - u)(b^2 - u)(a^2 - u)}} \, du$$

$$y_{proj} = \int_{c^2}^{v_i} \sqrt{\frac{-v}{(c^2 - v)(b^2 - v)(a^2 - v)}} \, dv.$$

(5)

The Integral to calculate the horizontal and vertical coordinates in the Jacobi projection is chosen by the fact that the projection axis X_{proj} is directed horizontally to the right, and the Y_{proj} axis vertically upwards.

Intersections of the ellipsoid surface with hyperboloids of one sheet establish on the ellipsoid a system of curves and for each curve a value of u is constant, but v varies from c^2 to b^2. Intersection of the ellipsoid surface with hyperboloids of two sheets derives a system of curves for which v is a constant value, but u varies from b^2 to a^2. Together, these systems define coordinate lines that are orthogonal to each other, i.e., an orthogonal system of curvilinear coordinates. Figure 2 shows the projection (geometric) of the coordinate lines of the elliptic systems to the plane XZ.

Figure 3 shows these coordinate lines on a Jacobi projection to the plane $X_{proj}Y_{proj}$ with a background of a cartographical grid. The spacing of the elliptical coordinate grid is constant on u and v, but the distances between the lines in the projections are not identical. We see this type of grid because of x_{proj} depends of u only and y_{proj} depends of v only but the relation is not linear.

Fig. 2 The projection of the coordinate lines of the elliptic systems to the plane XZ
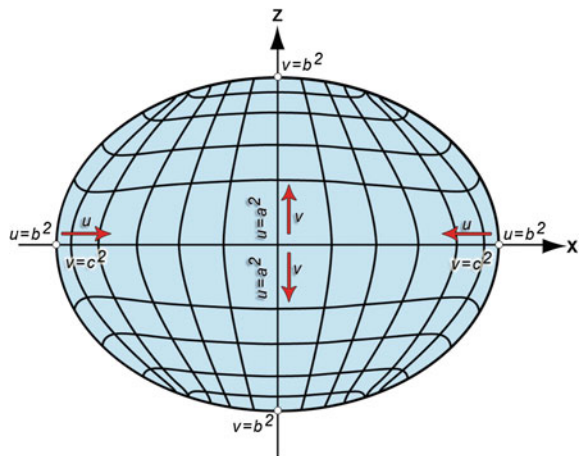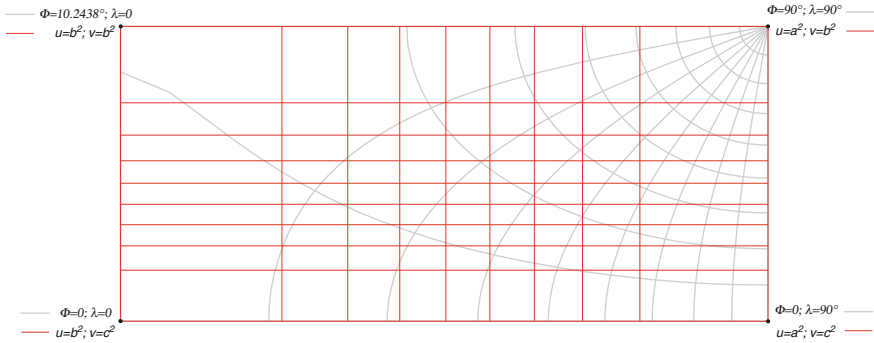
Fig. 3 Coordinate lines of elliptical system on Jacobi projection

Because the squares of three-dimensional rectangular coordinate values are used in the derivation of the formulae we obtain the values of the elliptic coordinates and rectangular coordinates in Jacobi projection respectively for one-eighth of the ellipsoid surface. For the other parts of the surface we flip coordinates in accordance with the sign of the latitude and longitude range.

4 Calculation of the Upper Limits of Integration

The upper limits of integration u_i, v_i in (5), we get for each point with the given latitude and longitude step by step.

1. Calculate three-dimensional rectangular coordinates as the function of latitude and longitude (geographical coordinates) See Fig. 1.

$$x = r \cos \phi \cos \lambda, \ y = r \cos \phi \sin \lambda, \ z = r \sin \phi. \tag{6}$$

Here:

$r = \frac{a}{\sqrt{t}}$, $t = \cos^2 \phi \cos^2 \lambda + \frac{\cos^2 \phi \sin^2 \lambda}{1-e_1^2} + \frac{\sin^2 \phi}{1-e^2}$, e_1—equatorial ellipse eccentricity, e—eccentricity of prime meridian ellipse.

2. Calculate elliptic coordinates as the function of three-dimensional rectangular coordinates, solving the Eq. (4).
3. Integrating (5), we obtain the rectangular coordinates in the projection.
   For the integration we perform a calculation based on Gaussian quadrature rule (GIS Research Centre Site 2012). This quadrature rule is a special case of the integral of the Lagrange interpolation polynomial with a special selection of nodes and weights. Direct calculation of integrals (5) is possible, but does not provide the required accuracy, therefore we express these integrals in terms of

elliptic integrals of the first and third kind, as it is done in (Prudnikov et al. 1986) Sect.1.2.35—formula 8 and Sect.1.2.36—formula 8.

$$x_{proj} = \int_{b^2}^{u_i} \sqrt{\frac{u}{(c^2 - u)(b^2 - u)(a^2 - u)}} \, du =$$

$$\frac{2}{\sqrt{(a^2 - c^2)b^2}} \left[(b^2 - c^2) I_3(\varphi_i, k_2, k_1) + c^2 I_1(\varphi_i, k_1) \right] \tag{7}$$

$$y_{proj} = \int_{c^2}^{v_i} \sqrt{\frac{-v}{(c^2 - v)(b^2 - v)(a^2 - v)}} \, dv =$$

$$\frac{2}{\sqrt{(a^2 - c^2)b^2}} \left[(b^2 - a^2) I_3\left(\frac{\pi}{2}, k_2, k_1\right) + a^2 I_1\left(\frac{\pi}{2}, k_1\right) \right] \tag{8}$$

$$- \frac{2}{\sqrt{(a^2 - c^2)b^2}} \left[(b^2 - a^2) I_3(\varphi_i, k_2, k_1) + a^2 I_1(\varphi_i, k) \right].$$

The calculation of the vertical coordinate (formula (8) of our paper) as the difference of the integrals due to the fact that in the formula 8 of Sect. 1.2.36 (Prudnikov et al. 1986) upper limit of integration is equal to the maximum value, and the lower one is the current value of the elliptic coordinate.

3.1. Express the integrals (5) to the elliptic integrals of the first and the third kind. Here φ is the variable of integration and k_1, k_2 are constant parameters for the calculation of integrals. φ_i is the upper limit of integration for elliptic integrals.

In integration of (7) $\varphi_i = \arcsin\sqrt{\frac{(a^2-c^2)(u-b^2)}{(a^2-b^2)(u-c^2)}}$, $k_1 = \sqrt{\frac{(a^2-b^2)c^2}{(a^2-c^2)b^2}}$, $k_2 = \frac{a^2-b^2}{a^2-c^2}$, and in integration of (8) $\varphi_i = \arcsin\sqrt{\frac{(a^2-c^2)(b^2-v)}{(b^2-c^2)(a^2-v)}}$, $k_1 = \sqrt{\frac{(b^2-c^2)a^2}{(a^2-c^2)b^2}}$, $k_2 = \frac{b^2-c^2}{a^2-c^2}$.

Elliptic integral of the fist kind:

$$I_1(\varphi_i, k_1) = \int_0^{\varphi_i} \frac{d\varphi}{\sqrt{1 - k_1^2 \sin^2 \varphi}}. \tag{9}$$

Elliptic integral of the third kind:

$$I_3(\varphi_i, k_2, k_1) = \int_0^{\varphi_i} \frac{d\varphi}{(1 - k_2 \sin^2 \varphi)\sqrt{1 - k_1^2 \sin^2 \varphi}}. \tag{10}$$

3.2. After substituting (9) and (10) to (7) and (8) we integrate given expressions and get coordinates of the Jacobi conformal projection in a range of latitudes from the equator to the North Pole and longitude from the prime meridian to the 90° meridian.

3.3. Flip coordinates in accordance with the actual position of point on the triaxial ellipsoid.

3.4. Multiply coordinates by the size of the major semi-axis in units of the map. In that case, when, $u = b^2$ and $v = b^2$ we get the so-called circular points (Kagan 1947, p. 117) of the triaxial ellipsoid. Taking into account that longitude is equal to zero, we obtain

$$t = \cos^2\phi + \frac{\sin^2\phi}{1 - e^2} = \frac{\cos^2\phi - e^2\cos^2\phi + \sin^2\phi}{1 - e^2} = \frac{1 - e^2\cos^2\phi}{1 - e^2},$$

$$x^2 = r^2\cos^2\phi = \frac{a^2\cos^2\phi\,(1 - e^2)}{1 - e^2\cos^2\phi}, \ i.e. \ \frac{a^2\cos^2\phi\,(1 - e^2)}{1 - e^2\cos^2\phi} = \frac{a^2(a^2 - b^2)}{a^2 - c^2}$$

$$\frac{\cos^2\phi\,(1 - e^2)}{1 - e^2\cos^2\phi} = \frac{e_1^2}{e^2}$$

$$e^2\cos^2\phi\,(1 - e^2) = e_1^2 - e^2 e_1^2\cos^2\phi$$

$$\cos^2\phi = \frac{e_1^2}{e^2\left(1 - e^2 + e_1^2\right)}.$$

5 Mapping

The projection exists at all points of the ellipsoid. Coordinate values of the projection are nowhere equal to infinity. However, conformality doesn't exist in circular points, and a gap in the projection is required for the prime meridian and its opposite or between the equator and the circular point, or between the circular point and a pole. Figure 4 shows the cartographical grid of a conformal Jacobi projection for a triaxial ellipsoid with the parameters $a = 267.5$ m, $b = 147$ m, $c = 104.5$ m, for the asteroid 25143 Itokawa, and parallels passing through the circular points. Latitude in degrees of the circular points at such parameters is $\pm 10.2438°$. A gap occurs on prime meridian (and 180° meridian) in the northern hemisphere from the equator to the circle point $+10.2438°$ and in the southern hemisphere from the circle point $-10.2438°$ to the equator.

Figure 5 shows a map of Itokawa asteroid in Jacobi conformal projection created by transformation of a photomosaic of the asteroid surface in the azimuthal equidistant projection of a sphere originally created by Philip Stooke. The local-
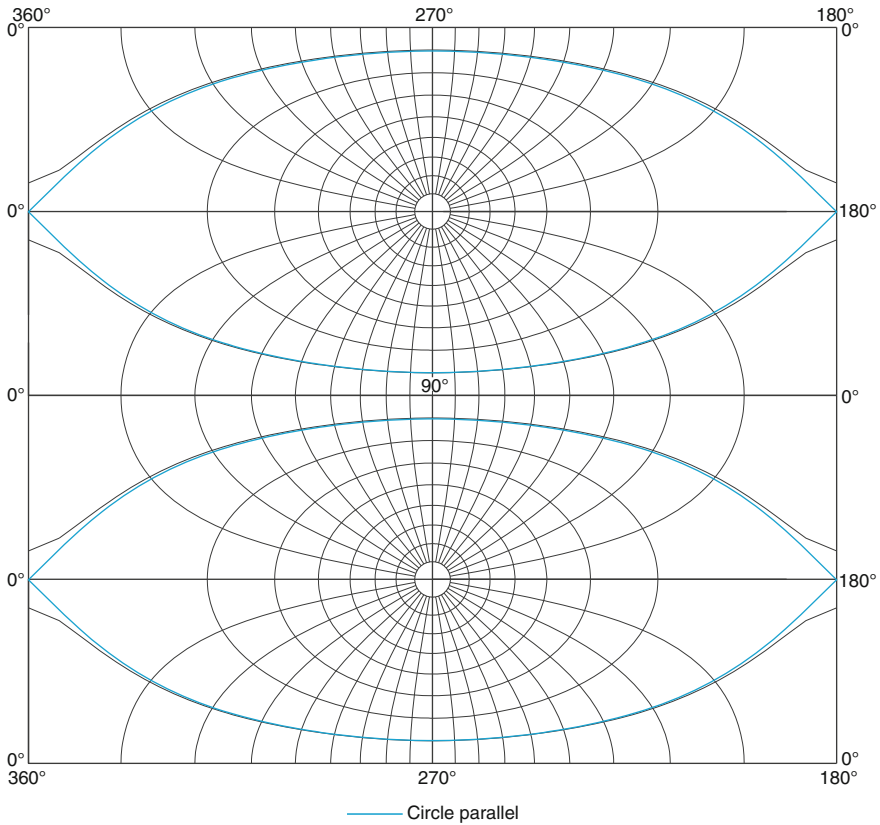
Fig. 4 Cartographic grid in Jacobi conformal projection

affine transformation was used in GeoGraph GIS 2.0 software. The control points for transformation were obtained by the tool for the calculation of the Jacobi projection (GIS Research Centre Site 2012). Unfortunately, due to an incorrect transformation in the neighborhood of circular points the map excluded areas located closer than 20° to the prime meridian and the meridian opposite it. A linear cartographic grid layer and a point craters layer with coordinates taken from (Gazetteer of Planetary Nomenclature—USGS 2012) were imposed on the resulting photomosaic.

Taking into account that the Jacobi projection is based on the relationship between semi-axes sizes a > b > c, in the special case when the polar flattening is equal to equatorial flattening, it is possible to use the cylindrical projection with right angle between meridian and parallel in the transverse orientation. The projection is also present in GIS Research Centre Site (2012). Figure 6 shows a map of asteroid 433 Eros in this projection. The new pole of the triaxial ellipsoid is the point of intersection of the equator and the prime meridian, and the major axis
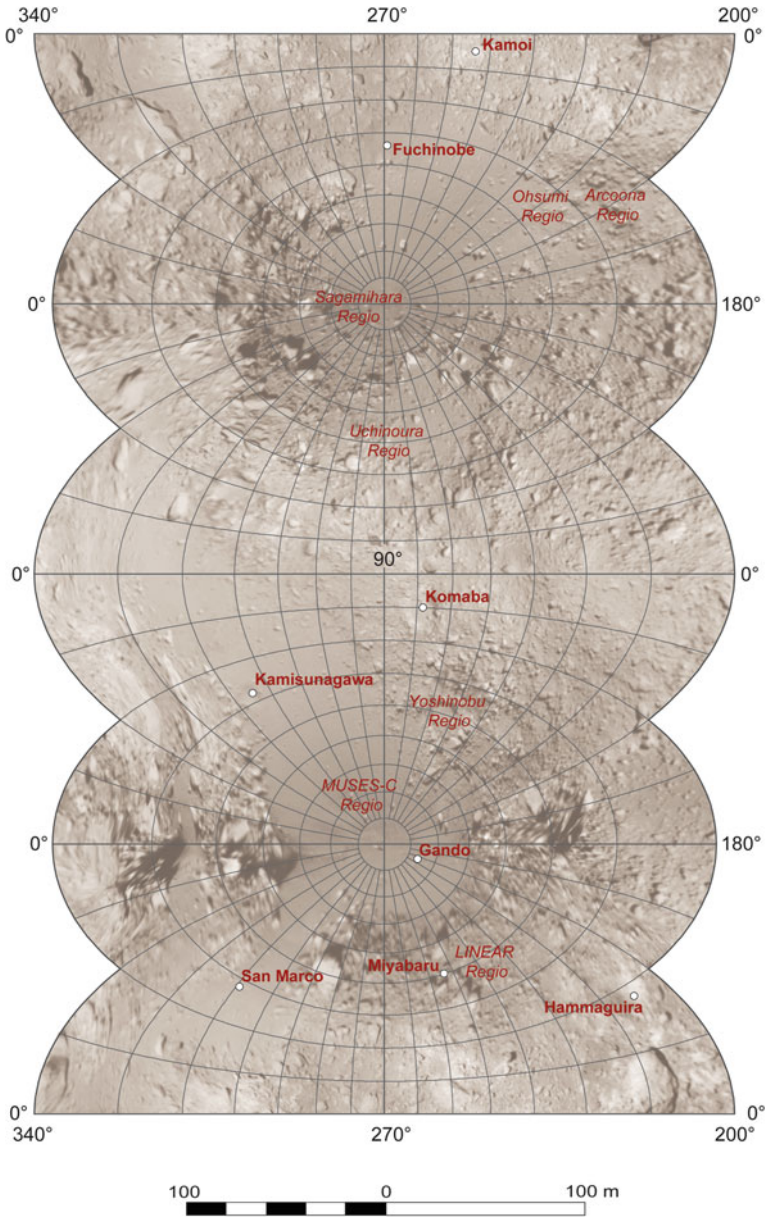
Fig. 5 Map of asteroid 25143 Itokawa in Jacobi conformal projection

becomes polar. In the case when the minor axis of the initial ellipsoid is equal to its polar axis, the resulting projection is conformal, as in fact it is a projection of the ellipsoid of revolution. A map was also created in GeoGraph GIS 2.0 software by transforming a base global photomosaic of the surface of an asteroid from the
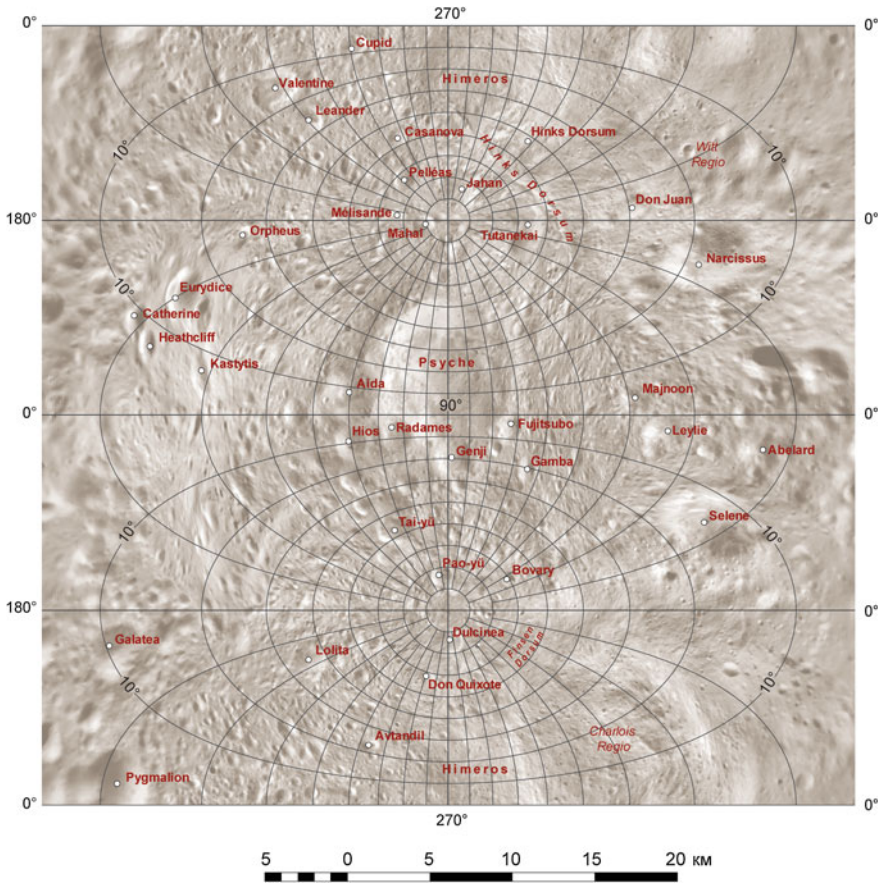
Fig. 6 Map of asteroid 433 Eros in the transverse conformal cylindrical projection of the triaxial ellipsoid

equidistant azimuthal projection of a sphere to transverse cylindrical projection of the triaxial ellipsoid. The transformation was made by special plug-into the module of calculating the projections on the local workplace. Therefore only a small neighborhood of points going to infinity was excluded from a map.

6 Conclusion

Cartographic grids in the transverse conformal cylindrical projection of the triaxial ellipsoid and in the Jacobi conformal projection have some similarities in the area of the pole and the neighborhood of central meridian. As a result, the technology for mapping in these projections also has similar particularities. We use a two step

process for transformation of photomosaics: from a simple cylindrical projection into azimuthal projection first and the transformation of obtained photomosaics to conformal projection of the triaxial ellipsoid after. There is also an important feature in common between the transverse conformal cylindrical projection of the triaxial ellipsoid and the Jacobi projection derivation. In both cases, we first go from the latitude/longitude coordinate system to the orthogonal curvilinear coordinates on the surface of the triaxial ellipsoid, and then to the mapping on the projection plane. Only for cylindrical projection the orthogonal coordinates are latitude and longitude referred to the poles and the equator of transverse system, and for the Jacobi projection those are elliptic coordinates.

References

Bugaevskiy LM (1999) Teoriya kartographicheskich proekciy regulyarnych poverchnostey. M. Zlatoust (in Russian)

Clebsch A (2009) (ed) Jacobi's lectures on dynamics, 2nd edn. Hindustan Book Agency, New Delhi

Gazetteer of Planetary Nomenclature—USGS (2012) http://planetarynames.wr.usgs.gov/. Accessed 26 Oct 2012

GIS Research Centre of the Institute of Geography of the Russian Academy of Sciences Cartographical Projections of Triaxial Ellipsoid (2012) http://geocnt.geonet.ru/en/3_axial. Accessed 26 Oct 2012

Jacobi CGJ (1866) Vorlesungen über Dynamik. G. Reimer, Berlin

Jakobi K (1936) Lekcii po dinamike/Perevod s nemeckogo O.A. Polosukhina/Pod redakciey Prof. N.S. Koshlyakova. Glavnaya redakciya obshetekhnicheskoy literatury. Leningrad-Moskva (In Russian)

Kagan VF (1947) Foundations of the theory of surfaces in a tensor setting, Moscow-Leningrad (In Russian)

Prudnikov AP, Brychkov YA, Marichev OI (1986) Integrals and series. Elementary functions, vol 1. Gordon and Breach, New York

Exploring Martian Climatologic Data Using Geovisualization: MARSIG a Spatio-Temporal Information System for Planetary Science

Paule-Annick Davoine, Christine Plumejeaud, Marlène Villanova-Oliver, Isaac Bareto, Pierre Beck, Bernad Schmit and Jérôme Gensel

Abstract The accumulation of observation data about the planet Mars during the last decade has contributed to the development of the Martian climatology field. This field focuses on the study of temporal evolution of physical proprieties of surface and atmosphere of the planet Mars. Taking into account the temporal dimension is a new topic in planetary sciences, which requires new methodologies and tools to explore data. Data about Mars come from Mars-Express or Mars Reconnaisance Orbiter spacecrafts. These data are multi-dimensional, including spatial, temporal, spectral and thematic components. They are also extremely heterogeneous and incomplete. To carry out their studies, the researchers need to identify and to extract some relevant dataset for a region of interest and a selected time period. This paper presents MARSIG, a spatio-temporal information system dedicated to explore and visualize Martian climatologic data. First, the main characteristics of the Martian climatologic data are presented, and the needs of the researchers in planetary sciences in terms of exploration and visualization are discussed. Then, we present how the different dimensions of these specific data,

P.-A. Davoine (✉) · C. Plumejeaud · M. Villanova-Oliver · I. Bareto · J. Gensel
Laboratoire d'Informatique de Grenoble, Saint-Martin-d'Hères, France
e-mail: Paule-Annick.Davoine@imag.fr

C. Plumejeaud
e-mail: christine.plumejeaud-perreau@univ-lr.fr

M. Villanova-Oliver
e-mail: Marlene.Villanova@imag.fr

C. Plumejeaud
UMR Littoral Environnement et Sociétés, La Rochelle 7266, France

P. Beck · B. Schmit
Insittut de Planétologie et d'Astrophysique de Grenoble, Université Joseph Fourier, CNRS, Saint-Martin-d'Hères, France
e-mail: Pierre.Beck@ujf-grenoble.fr

B. Schmit
e-mail: Pierre.Beck@ujf-grenoble.fr

and more especially temporal dimension, have been integrated into a geovisualisation interface to answer their needs.

Keywords Martian climatologic · Geovisualisation · GIS planetary

1 Context and Objectives

The Martian climatology science addresses the spatio-temporal evolution of the physical proprieties of the surface and atmosphere of the planet Mars. For example, it concerns the study of the localization of some chemical components at the surface of Mars, but also the seasonal variations of these components. The aim is to learn more about the geographical coverage of various components, like CO_2 that, for example, is involved the ice formation.

To carry out these investigations, the Martian Climatologic science uses large spatial data coming from satellite remote sensors, especially the scene from Mars-Express or Mars Reconnaissance Orbiter spacecrafts (Esposito et al. 2007). These spacecrafts, in orbit on Mars, are equipped with sensors or cameras to measure the level of light reflectance from the surface of the planet. Following their trajectories (or orbits) they make over flights at different dates, during which they perform sets of various "images". Each image is multidimensional, made of irregular pixels containing reflectance information for various wavelengths, and includes more than 20 Mo of information.

The various chemical components, such as water, CO_2 or dust, do not reflect the same wavelengths. The sensors record the level of soil reflection of 256 wavelengths. Therefore, the processing of sensors data allows for checking, according the value of the albedo, the existence (or not) of specific chemical components. Each pixel in the image is associated with a spectral pass band, and consequently with a chemical profile. Then, the analysis of this spectral pass band leads to the characterization of the physical properties of the Martian surface (Christensen 1988).

These data are then used with the objective to map the Martian Polar ice cap. The idea is to map areas where the water's presence has been detected, or areas mainly with dust, etc. Another objective is also to produce energy balances on the long term, relying on the fact that the CO_2's presence is a sign for the formation of ice and for the changes of this energy balance as well.

Before producing such maps, it is essential for researchers in planetary climatology to get an inventory of available data and to explore them according spatial, temporal and spectral dimensions. This exploration not only allows analysing the characteristics of the Martian surface, but also makes it possible to perform a comparative study of these characteristics both at the whole planet scale and over time. However, data samples are not homogeneous. Sample quality is very variable, depending on the observation conditions. Before any further

investigation, it is thus necessary to identify and select a relevant dataset characterizing a specific geographical area called *Region Of Interest (ROI)*. Two criteria are important for this: the density of data concerning the selected ROI and the quality of the information contained into the datasets. From a methodological point of view, this involves:

- Finding among all the data produced, subsets of pixels covering a given ROI, for a selected spectral range and selected time period.
- Identifying geographical areas having a maximal or minimal density of information over time or identifying temporal periods where is possible the observation of given frequencies of the albedo on Mars surface.

Exploring and identifying available data are essential steps in the study of the characteristics of the Martian surface as performed by planetary researchers. Martian climatologic data being spatio-temporal data, these steps can benefit from methods and tools for geo-visualization and exploratory data analysis (Andrienko et al. 2006). These two observations are the first conclusions drawn by a multidisciplinary consortium involving the Planetology and Astrophysics Institute of Grenoble (IPAG) and the Grenoble Informatics Laboratory (LIG). Together, a team specialized in the analysis of hyperspectral data (Planeto-IPAG), and a geomatics team (Steamer-LIG) have designed and developed a spatio-temporal information system for the exploratory analysis of planetary data coming from spectro-imagers embedded in Martian orbiters: OMEGA (Mars Express) et CRISM (MRO). This software is called MARSIG.

The paper is organized as follows. The second section presents the specific difficulties that may arise when analysing and exploring hyperspectral data. The approach conducted by planetary researchers is also described. The third section explains how the MARSIG software takes into account these difficulties and integrates an exploratory approach to analyse planetary information.

2 Martian Climatologic Data

Recent advances in data acquisition in the field of planetary observation, especially those regarding Mars, have produced numerous sets of records, having spatial and temporal attributes. Large volumes of multidimensional data like raster images, and derivatives (raster or vector) are to be organized, structured and visualised for their further exploitation. We detail here the research process and the specificities of raw raster data that made necessary the design and development of a new tool for their exploration and understanding.

2.1 Martian Products

Climatologists and glaciologists manipulate two kinds of data: raw data and derived ones. Raw data are collected first, under the form of images or telescopic observations on various wavelengths. From these raw data new products are elaborated. These new products result from the expertise of physicians or chemists, relying on the astrophysics science in general. For instance, the presence of H_2O and CO_2 molecules is monitored, either by the analysis of albedo based on raw raster data, which can produced polylines of isolvalues of physical parameters like CO_2 density, or by interpolating the raw punctual samplings of the ground made by Spiders, which gives density surfaces. Building derived products is never immediate since it requires the use of sophisticated mathematical artefacts and computational resources.

The new products often take the form of vector data, such as for instance a map of physical parameters of the Martian surface describing mineral water concentration, different kinds of ice that can be found, or any other kind of chemical component of interest for the study, etc. represented by polygonal areas. These derived products (e.g. a map of permafrost characteristics: depth, thickness, abundance of ice), are issued from a research process consisting in mining raster images, by analysing the albedo for instance, in order to detect and cartography the presence of certain chemical component. The process makes also a large use of additional data, raster or vector ones, consisting of existing maps of Mars, for instance, and all previous collected knowledge on this planet. It can be topologic data under the form of a Digital Elevation Model (DEM), and geologic maps (raster or vector forms as well).

One of the objective is to build 3D maps allowing for the modelling and representation of sedimentary stratification and stratification of subsurface ice. The second objective is to follow the evolution of this area, by studying for instance the seasonal variations of polar caps, soil hydration, and cratering of the ground.

At this stage, it is important to understand that the process is iterative, because it re-uses previous derived products and additional data in order to enhance the interpretation of the new raw data, or even to fix possible misinterpretations of previous raw data, and correct the scientific principles and theory that are guiding the scientists. This means that the exploration of any new raw data, the hyperspectral cube as it will be described in the next section, needs to be combined with all previously acquired heterogeneous datasets (DEM, polygonal areas, punctual samplings, raster data, etc.).

2.2 Constitution of a Hyperspectral Cube

Researchers manipulate data file provided by satellites images capturing the albedo of various places on Mars (see Fig. 1). With a size of 20 MB, such a file contains X * Y pixels of different shapes and irregular surfaces. For instance, referring to

Fig. 1 Schema of a satellite taking a series of pictures on orbit
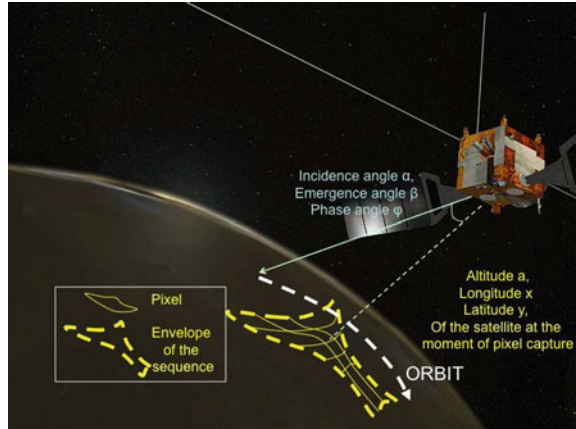
Fig. 1, the corresponding data file is constituted of 5×3 pixels forming together the spatial envelop of the sequence (bold yellow line) that represent the full trace scanning by the satellite following an orbit at a certain date. Each pixel is characterized by the position of the satellite at the time of image acquisition (angles of incidence, emergence and phase, longitude, latitude and altitude measured in Martian coordinates system). A pixel is composed of 256 photos, one for each wavelength. Each pixel is distinctively identified by its serial number (order of taking) in a given sequence for given orbit and time. Pixels share some properties like the resolution and the sampling size of the satellite sensor. The spatio-temporal extent of the sequence can be computed from the pixels features.

Therefore, in a more abstract way, pixels form the so-called hyperspectral cubes (Fig. 2) which are the most difficult raw data to analyse.

Only one pixel at once can be analysed by software like ENVI[1] for instance. Indeed, it takes times to load it into memory, and also to run the appropriate tool in order to extract albedo profiles for each wavelength. The main problem for planetologists is to identify the pixels that are "interesting", that is to say of good quality and of sufficient density. The first problem comes from the disparate quality of pixels, making some of them be almost unusable for further analysis following:

- the position of the satellite in the space (the angles of incidence α and emergence β should not be too small),
- the brightness of the ground related to the phase angle ϕ and the position on LS,
- and the location of a pixel inside a sequence (pixels have a tendency to be more thick and distorted on the border of the sequence).

Thus, volume and number of pixels contained in hyperspectral cubes (up to 1000 sequences per year), as well as their heterogeneity in shape and quality make them difficult to explore.

---

[1] http://www.exelisvis.com/ProductsServices/ENVI/ENVI.aspx.

Fig. 2 Hyperspectral cube: sequence of X * Y on 256 wavelengths
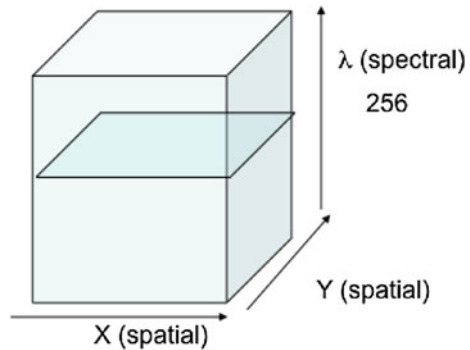
Also, since scientists are interested in evolution studies, they would like to get pixels covering the same zones at various time intervals. However, sequences of satellite orbits flights rarely strictly overlap. Here the temporal dimension of these data must be well understood, since density of pixels covering the same zone may not have the same interpretation following the season of the picture.

The temporal dimension of information is associated with the date of satellite flight (absolute time), but also with the position of Mars in its solar orbit (Solar Longitude—LS) that will determine the season of picture taking, as shown on (Fig. 3). Indeed, depending on the season, and therefore on more or less exposure to high-energy solar radiation, results in terms of density of water or ice should be interpreted differently by glaciologists. Let us consider an existing series of images taken on almost identical orbits over several years, measured along a Martian calendar,[2] which started a long time ago after the Gregorian calendar used in the ISO 8601 standard. In this series, one should be able to detect sequences covering the same temporal period, a given season for instance. But also, using a cyclic time model, based on Martian seasons, one should help scientist to explore derived data, such as abundance of ice maps, grain size of ice maps, etc. The objective is to compare products between them in order to draw correlations, find seasonal tendencies or long-term trends.

The variables, which correspond to metadata, describing a hyperspectral cube to consider are then:

- Number of order: O (fixe)
- Orbit number: On(float)
- Sequence number: Sn (int)
- Solar longitude: LS (float)
- Martian year: Y(int)
- Absolute time: T (float)
- Column number: Y (int)
- Line number: X (int)

---

[2] Here, one year corresponds to 600 earthling days.

Fig. 3 Seasons on Mars,
depending of LS
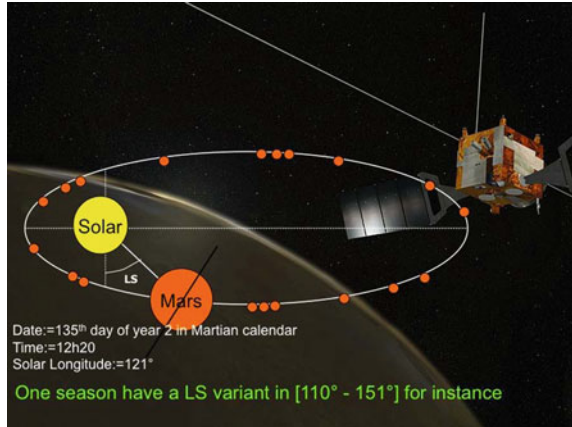
Table 1 Martian climatologic data: some spatial characteristics

Spatial extension corresponding to	Spatial resolution	Spatial sample
The light areas at time T,	Variable according the images, but also according the period	Very variable, generally decreases with the latitude
		Highly dependent on the location
The polar ice caps, (North and South hemispheres)	Variable within an image very different between OMEGA and CRISM	Sometimes only a few traces or series allowing to create a local or regional map

- Maximum and minimum Latitude: MaxLat, MinLat
- Maximum and minimum Longitude: MaxLong, MinLong Concerning the spatial dimension of matian climatologic data, differents characteristics may be take into account (Table 1).

The georeferencing is generally known, but sometimes it is inaccurate. It is based on polar stereographical projection (N and S), or equatorial projection (cylindric).

3 Toward a Spatio-Temporal Information System for Martien Climatology

3.1 An Exploration Tool

The use of GIS in the field of planetary science has developed significantly during several years (Hare et al. 1998, 2003; Tanaka and Kolb 2001; Deuchler et al. 2004). The main application of GIS tools concerns mapping the surface of Mars.

However, the integration of planetary products into a GIS arises various problems related to:

- the heterogeneity and the variable quality of spatial samples. The spatial extension of the planetary products is limited to lighted area at time T, and is not homogenous over time;
- the specificity of the temporal dimension of these planetary products. Indeed, they are defined according several temporal granularity levels (daily, seasonal, annual or pluri-annual scales);
- the spatial resolution, which can be different between the images, but also within a same image;
- the diversity and specificity of projections (cylindric or polar projections, …) that have to be combined;
- the amount of data to explore which can reach up to several terabytes;
- the multi dimensional nature of data (Fig. 3):

  - informational or descriptive data: Orbit number, sequence number, file name;
  - temporal data: date of capture (LS1 and LS2, Martian year);
  - spectral data: 256 wavelengths or plans;
  - spatial data: pixel, spatial envelope.

Martian planetary research is characterized by an exploratory approach, which complicates the information system design and implementation. In fact, needs in terms of query are not identified a priori since the work is highly exploratory, consisting in highlighting geographical areas (region of interest) where the temporal and spatial coverage are of good quality, and having an interesting spectral signature (presence of ice, CO_2 and H_2O, not much of atmospheric dust that disturbs the signal…). This exploration may also involve the visualization of variety of geographical products resulting from Mars studies (raster or vector). Once the region of interest of study has been identified, the corresponding spectra can be extracted and processed by other softwares, in order to identify the physical properties of Martian surface.

3.2 Main Functionalities

From the analysis of needs, we have identified the main functionalities that the information system should provide. These needs are:

- to explore data files in order to identify regions of interest according to some criteria like density or diversity of data (main input criteria in the process analysis performed by planetary researchers);
- to extract data corresponding to the selected region;
- to realize maps and graphical representations of physical properties of the surface of Mars. These representations are dedicated to the understanding of

Mars meteorology and climatology, and to the localization of the quantities of water.

In a more concrete way, the tool should allow researchers:

- to visualize simultaneously and interactively in space and in time, the various "properties" produced by OMEGA and CRISM: raster products (maps) and vector products (polygons and polylines). This visualization is guided by the spatial axe i.e. "latitude", either for a chosen latitude;
- to visualize spatially temporal phenomena;
- to compare spatial or spatio-temporal products with each other;
- to identify spatial areas with good temporal datasets available for a given temporal period (homogeneous sample);
- to extract and to visualize these ROI, and the temporal evolution of their properties;
- to extract, for any selected point, the temporal evolution of various properties (raster and vector);
- to perform spatio-temporal analysis based on vector processing (topological queries…);
- to perform attribute queries on parameters of products (date, time…);
- to manage large datasets (image's size from 100 Mpix up to 1 Gpix).

Table 2 shows some examples of expected spatio-temporal queries:

Queries and exploration should be considered at different levels of granularity in time: annual, seasonal and inter-annual.

In addition to the features of exploration and visualization, the proposed information system should support interoperability with softwares commonly used by the planetary researchers to conduct theirs analysis, and to create new products (raster or vector). Figure 4 shows the functional architecture of MARSIG Information system.

Table 2 Example of spatio-temporal queries

Identification of intersections or exclusions between regions of interest (polygons)	Identification of the evolution of region	Inter-annual comparison about area	Thematic
Zones containing H_2O and CO_2, or zones containing H_2O but no CO_2	Are the boundaries of the cap for T1 and T2 periods the same? Is polygon for T1 included into polygon for T2 ?	Difference between the ice region extension for two periods	Discover the presence of some chemical components on Mars surface by spatial (where?), and temporal queries (when?)
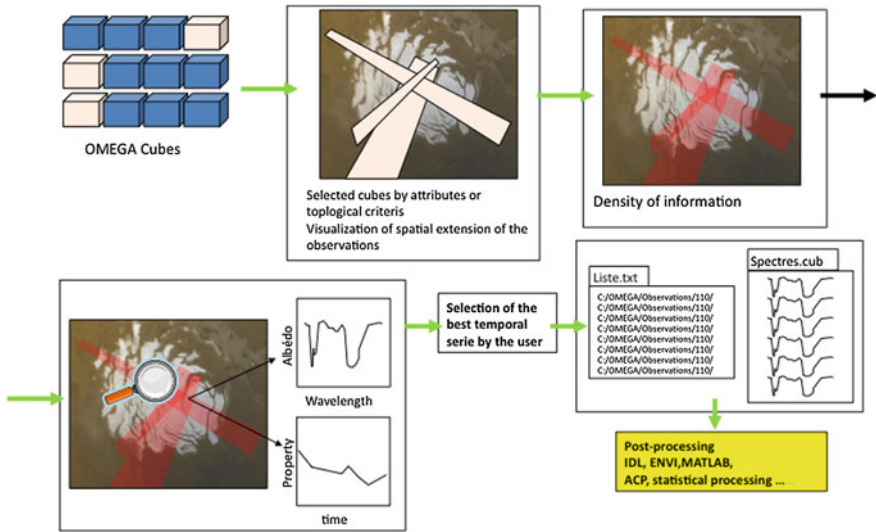
Fig. 4 Functional schema of MARSIG tool

3.3 Geovisualisation Interface

MARSIG is a geovisualisation tool for Mars planetary multidimensional data. It allows the exploration of the features of Martian surface areas that have been photographed by spacecraft at a certain time. These images have a spatial footprint, which is defined by dated polygon and associated with a data cube that contains descriptive and spectral data. MARSIG allows to index these data cubes, both by spatial footprint and by temporal footprint, but also to query them by *ad hoc* spatial and temporal criteria (multiple temporal granularity according the Martian Calendar, cylindrical or stereo-polar projections, …).

MARSIG is composed of several modules:

- An administration module, responsible for managing the data needed to run the search module (raw or elaborate products and images). This module also imports geographic data contained in ESRI shape files (files containing polygons), together with their metadata.
- A search module whose aim is to identify the polygons corresponding to the spatial and temporal queries of the user. These queries can be either visual or textual.
- A module in charge of displaying the selected products on the planet Mars map.

Thus, MARSIG offers to the user the ability:

1. View a study area of the surface of Mars (Central, North, South) and select the displayed area, some spatial subsets (or areas of interest) using a visual query (selection by clicking a point or drawing a bounding box on the map) or by entering coordinates in a text field (Fig. 5).
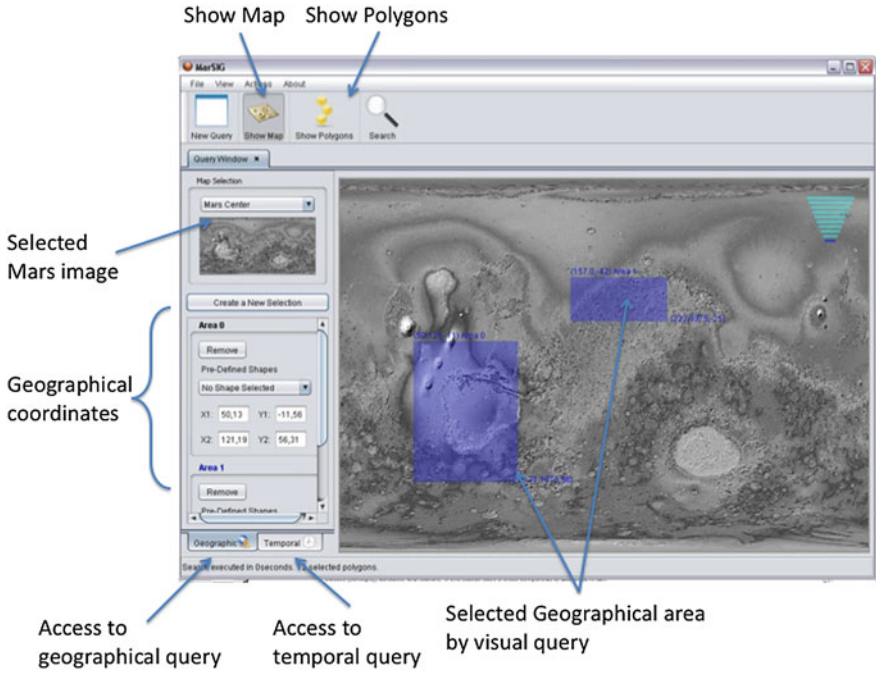
Fig. 5 MARSIG interface: spatio-temporal query

2. Specify criteria with different temporal granularities (annual, seasonal, daily, …) using graphical or textual components (Fig. 6).
3. To visualize polygons corresponding to cubes covering the selected region of interest (Fig. 7a, b).
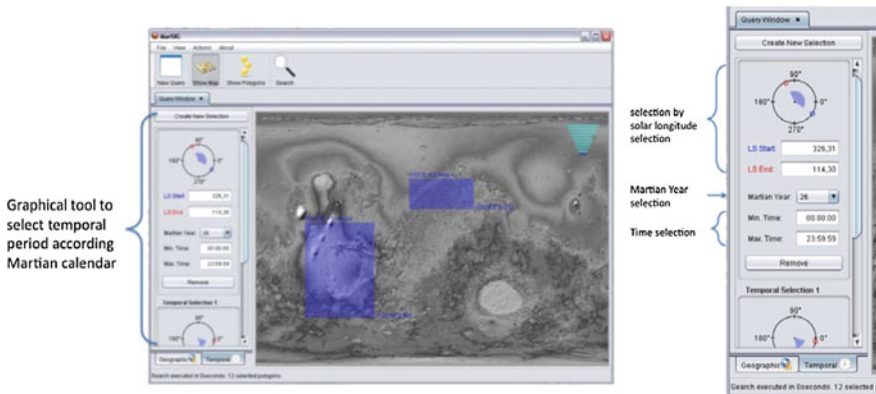
Fig. 6 MARSIG interface: temporal query
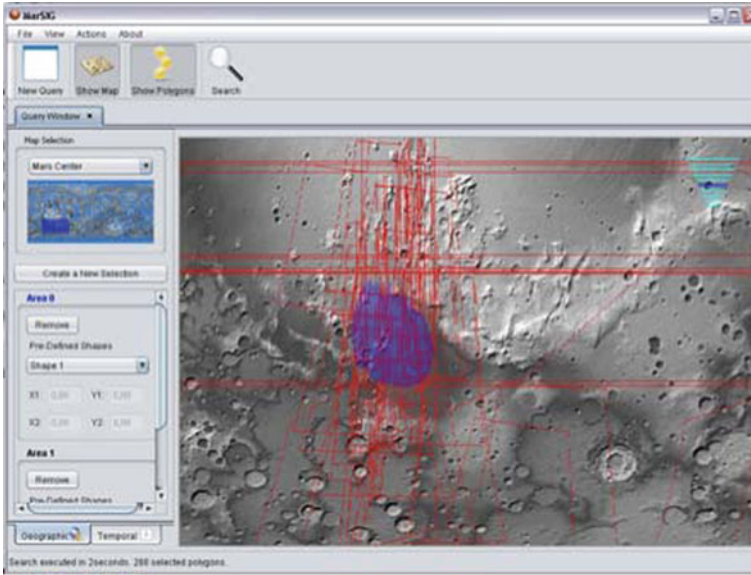
(a)
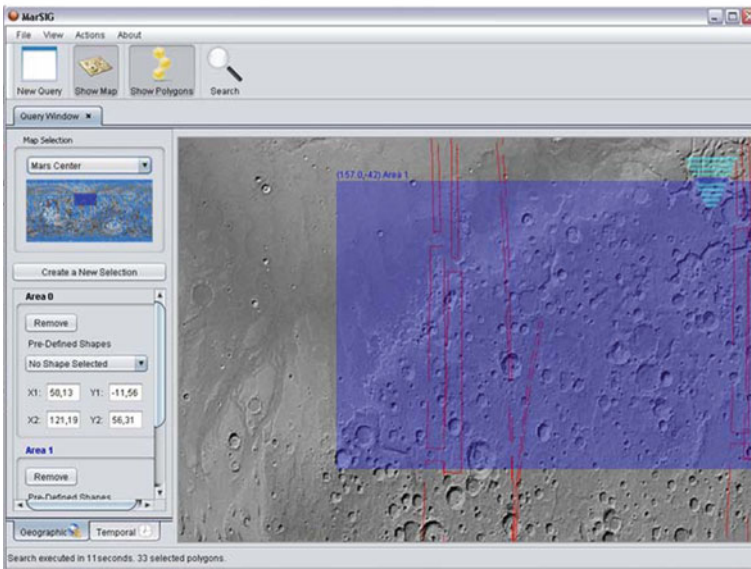
(b)

Fig. 7 a Interface MARSIG: Displaying polygons for a selected area of interest. **b** MARSIG interface: zoom in on some selected polygon

4. Elaborate spatio-temporal queries on database files containing polygons (shape file format) of the planetary products and planetary, and these products as map layers.

3.4 Technological Aspects

MARSIG is a desktop tool whose different modules have been developed in Java (Java Swing for the development of graphical components, and the display of the images). The data are storing into a PostGres/PostGis database.

As the data layers are extremely large (20 Mo for each image, and 100 000 images by year), we have chosen to process metadata files of images in order to make temporal and spatial indexation of these images. These technical choice had lead us to develop a script in order to create a shapefile which describe the convex envelop of the images. This script must be run for each integration of new data. Since updates are not so frequent, this technical proposal is efficient at long term, in terms of both storage and computing resources.

4 Conclusion

In this paper, we have presented the result of an interdisciplinary collaboration between geomatics sciences and planetary sciences, in order to get an efficient open source software as support to a scientific methodological approach for hyperspectral analysis.

After describing the specificity of Martian climatologic data, and the various treatments carried out these data, we have identified the way planetary science researchers exploit these data and defined theirs needs in terms of functionalities. A software called MARSIG for exploration and geovisualisation hyperspectral data is proposed.

MARSIG offers the ability to explore quickly very large datasets by exploiting the metadata embedded into the images. Using metadata, we propose a tri-dimensional interface to query and to display data according to the dimensions "What, Where, When". In particular, the temporal dimension as it exists in the Martian calendar has both a linear and a cyclic expression. This temporal specificity has led us to innovate in terms of query interface on this dimension.

For future research orientations, and in order to make a complete integration of MARSIG into the analysis process of the planetary researchers, it seems important to link this tool with hyperspectral data analysis open source softwares. We also investigate the pairing of MARSIG with other tools related to the analysis or the exploration of data quality (QualESTIM, Plumjeaud et al. 2012), which suggests new challenges, due to the specificity of these multidimensional and big data.

References

Andrienko N, Andrienko G (2006) Exploratory analysis of spatial and temporal data. Springer, Berlin

Christensen PR (1988) Global albedo variation on Mars—implications for active aeolian transport, deposition, and erosion. J Geophys Res 93(B7):7611–7624

Deuchler C, Wählisch M, Gehrke S, Hauber E, Oberst J, Jaumann R (2004) Combining Mars data in GRASS GIS for geological mapping, vol xxxv. In: IAPRS, Istanbul

Esposito F, Giurannab M, Maturilli A, Palomba CL, Formisano V (2007) Albedo and photometric study of Mars with the planetary Fourier spectrometer on-board the Mars express mission. ScienceDirect Icarus 186:527–546

Hare TM, Dohm JM, Tanaka KL (1998) GIS and its application to planetary research. Lunar Planet Sci XXVIII. http://webgis.wr.usgs.gov

Hare TM, Tanaka KL, Skinner JA (2003) GIS 101 for planetary research, Technical report. http://webgis.wr.usgs.gov/pigwad/publications/Hare_isprs_mar03pdf

Plumejeaud C, Villanova-Oliver M, Gensel J (2012) QualESTIM: interactive quality assessment of socio-economic data using outlier detection. In: Brindging the geographic information sciences, proceedings of international AGILE'2012 conference, pp 143–160

Tanaka KL, Kolb EJ (2001) Geologic history of the polar regions of Mars based on Mars global surveyor data. I. Noachian Hesperian Periods Icarus 154:3–21

A Framework for Planetary Geologic Mapping

Andrea Naß and Stephan van Gasselt

Abstract Archives of published planetary maps hosted at the United States Geological Survey or other facilities consist of a large number of small to large-scale geologic maps of terrestrial planets, in particular the Moon and Mars. Along with recent and upcoming missions also to Mercury, the Outer Solar System moons, and asteroids systematic mapping of surfaces has received new impulses. As planetary geologic mapping today is performed by individual scientists not only in the US but also in Europe with dedicates mission programs and participations (ESA Mars Express, ESAJUICE, …) a general framework of mapping and in particular for organizing cartographic output is paramount. This work presented here provides a general overview of cartographic and data requirements in the context of collaborative mapping programs and establishes an innovative data framework that allows data integration, management and access in order to support communication of scientific results across disciplines and the public.

Keywords Mars · Geology · Cartographic framework · Collaborative mapping

A. Naß (✉)
Department of Planetary Geology, German Aerospace Center (DLR), Institute for Planetary Research, Berlin, Germany
e-mail: andrea.nass@dlr.de

A. Naß
Geoinformation Science Research Group, Institute of Geography, University of Potsdam, 14476 Potsdam, Germany

S. van Gasselt
Institute of Geological Sciences, Freie Universitaet Berlin, 14195 Berlin, Germany

1 Introduction and Background

Image-and special-purpose (thematic) maps from an essential element of planetary exploration as cartographic products provide coherently distilled and targeted pieces of information and a basis for knowledge extraction from a large amount of obtained data in the course of a mission lifetime. In order to achieve such a goal, maps and cartographic communication tools, such as colors and symbols need to be well-defined, i.e. standardized.

Planetary geologic maps have been receiving particular interest for laymen as well as for researchers from different disciplines due to their esthetically pleasing appearance and their detailed level of geoscientific information and complexity. Such maps represent highly condensed information with a high potential of deriving new knowledge based on the cartographic representation of a unit's distribution, its compositional attributes, its inherent temporal context, and the possibilities to geometrically reconstruct three-dimensional bodies from a two-dimensional sheet of paper. In combination, well-designed geologic maps provide means to reconstruct the four-dimensional history of a surface.

Currently agency-financed and survey-coordinated mapping programs are only substantiated in the US under the auspices of *NASA's Planetary Cartography and Geologic Mapping Working Group* (PCGMWG) funded through the *Planetary Geology and Geophysics Program*. Recent mapping endeavors coordinated by the *United States Geological Survey* (USGS) include local to regional mapping of Mars (1:200k–1:5M, e.g., Bleamaster et al. 2010), Venus (1:5M–1:10M, e.g., Chapman 1999), Moon (1:2.5M, e.g., Wilhelms and McCauley 1971), and the Galilean satellites (e.g., Williams et al. 2011), and new proposals are being designed for mapping of Mercury based on recent data from, e.g., NASA's Messenger. Until today, the USGS stored in the digital and analog archives several hundred published maps (USGS 2012).

Europe has not yet succeeded to establish a planetary geologic mapping program despite several attempts via the European Commission (Frame-work 7) and/ or supported by national geological surveys and the European Space Agency. Since the 1990s, Europe has become active in planetary exploration within spacecraft contributions (Mars Express, Venus Express, Huygensprobe, Exo Mars, Bepi Colombo, JUICE) and employment of dedicated mapping instruments, such as the High Resolution Stereo Camera onboard Mars Express (Neukum et al. 2004, 2009; Jaumann et al. 2007). Such instruments are providing terabytes of data ready to use for systematic planetary mapping. Apart from dedicated mission programs and participation of ESA member-state researchers, active involvements in international mission collaborations, such as in NASA's DAWN mission, also provide means for cooperative systematic mapping.

The focus of this work is put on the technical cartographic means for conducting such a collaborative (not concurrent) and distributed systematic mapping process. As such endeavors are usually coordinated by few and conducted by researchers spread over different locations, a common model is required which

allows to store, combine(!), access, update, revise and, finally, visualize data. Such models are usually employed using a relational database structure employed within modern GIsystems. Once established and once the scheme has been distributed among researchers and mappers, the highest possible level of homogenization is reached (van Gasselt and Nass 2011a; Nass et al. 2011a).

In particular we want to address the following main issues:

1. What is needed for GIS-based systematic mapping and what are the requirements for establishing such a framework?
2. How should communication and workflows of mappers be organized?
3. How can research and mapping results be communicated in the context of science and map data dissemination?

Answers to these low-level questions help to provide a solid basis for a common understanding of technical requirements and needs in the context of modern GIS-based mapping. In particular they contribute to sets of higher-level specifications and allow establishing basic standards.

2 Planetary Mapping and Cartography

Planetary discovery and exploration started with first observations of the Moon and planetary objects that were reachable with the unaided eye. In 1608 GALILEI started Moon observations with the help of his telescope which initiated the phase of telescopic planetary exploration. A large number of map sheets followed by individual observers. The cartographic era of planetary mapping begun little later with *Selenographia* created by Hevelius in 1647, and the *Mappa Selenographica* which was developed by Beer and Maedler and published in 1834 (Batson et al. 1990). Since that time, the knowledge of the celestial bodies has been growing steadily. Especially technological developments made in the last 60 years have played a key role. Sensor technology and computerized data processing have means for highly accurate exploration. Since initiation of systematic mission-based planetary exploration the diversity and amount of data have exponentially increased. To understand, not *only observe*, a planetary surface mapping is one of the most suitable tools. In planetary sciences, the term *mapping* is frequently being used as equivalent to *cartography*. However, the two terms describe different things as *cartography* is "*art, science and technology of making and using maps*" (ICA 2003) while *mapping* describes the effective manual methodology applied to create a map. Beside this, the term *planetary mapping* is also frequently used equivalently to performing systematic observations from spacecraft. Thus, aims for *planetary mapping* are either technical with respect to engineering and mission design or they are scientific with respect to fundamental research questions or they are cartographical with respect to designing planetary maps. Comparable to earth-based cartography, planetary maps may have different meaning related to (1) photography and digital raster imagery using compound satellite data and map

labels (e.g., Roatsch et al. 2006, 2012), (2) topography (e.g., Buchroithner 1999; Shingareva and Krasnopevtseva 2001; Shingareva et al. 2002, 2003; Planetologia 2012), (3) thematic aspects (e.g., USGS 2012; Lehmann 2006) including investigation area or landing site studies.

Comparable with terrestrial mapping conduct there are basic technical aspects affecting both, planetary mapping and cartography: spatial reference and cartographic coordinates have to be defined (e.g., Seidelmann 2005); naming has to be uniform (e.g., Hargitai 2006; IAU 2012), and in particular, digital reproduction techniques and analyzing tools (GIS-based and web-based mapping) have to be improved. Limitations and demands for improvements are subject of this work.

3 A Mapping Framework

3.1 Optimized GIS-Based Mapping Process

Geologic and geomorphologic mapping of planetary surfaces represents a highly specialized method for the visualization of spatial data within a scientific context. This method and resulting maps, however, are not based upon an *objective* approach by simply depicting thematic issues but it is based on a highly *subjective* process of identifying, interpreting and delineating mixed-type surface units and structures as spatial entities. In order to map these entities the process of map visualization is subdivided into four processing steps commonly known as (1) *data acquisition*, (2) *filtering*, (3) *mapping* and (4) *rendering* (Haber 1990; Wood 1996; Carpendale 2003).

In *systematic mapping programs* (SMP) these four tasks are usually controlled by a number of different actors concerned with data processing and information retrieval techniques, basic research questions, and cartography on different levels (van Gasselt and Nass 2011a, b). Outside of such SMP context these tasks usually have to be performed by a single person, mostly a researcher who deals with the analysis and interpretation of science data for publication reasons. Consequently, the cartographic framework that is usually employed on an ad-hoc basis is highly limited and technical and scientific problems occur during mapping conduct, and, more specifically, when access to primary map data is needed in the future.

Geologic maps usually depict surface materials and an associated time frame. Each geologic-map unit therefore forms the smallest key element of any geologic map and it is a discrete entity of a measureable geologic feature. It is represented by its geometry, its relationships to and intersections with other units and geologic features, i.e. their topologic description, and by additional attributes. The detailed geoscientific means and approaches for delineating and classifying surface units are highlighted in, e.g., Wilhelms (1987, 1990) and Tanaka et al. (1994). For mapping conduct, modern GIS have become well-established tools and all steps required to obtain geologic maps are based upon knowledge and techniques that

were elaborated in an earth-oriented context. In order to make use of modern mapping, GI mapping systems need to be adapted and a mapping process needs to be formalized for the requirements of planetary mapping. Adaption of tools and formalization/specifications of mapping tasks are the basic steps in providing a framework for collaborative mapping. Such adaptions are targeted to:

(a) allow project and map design, storage and administration in a homogeneous and well-structured way, (b) allow updates and merging of project data, and (c) allow archiving for sustained use and for providing a primary data basis for subsequent mapping tasks.

For map data standardization and accessible storage of data, a software-based mapping framework must cover three major aspects: entity structure, visualization and attribute description (see Fig. 1).
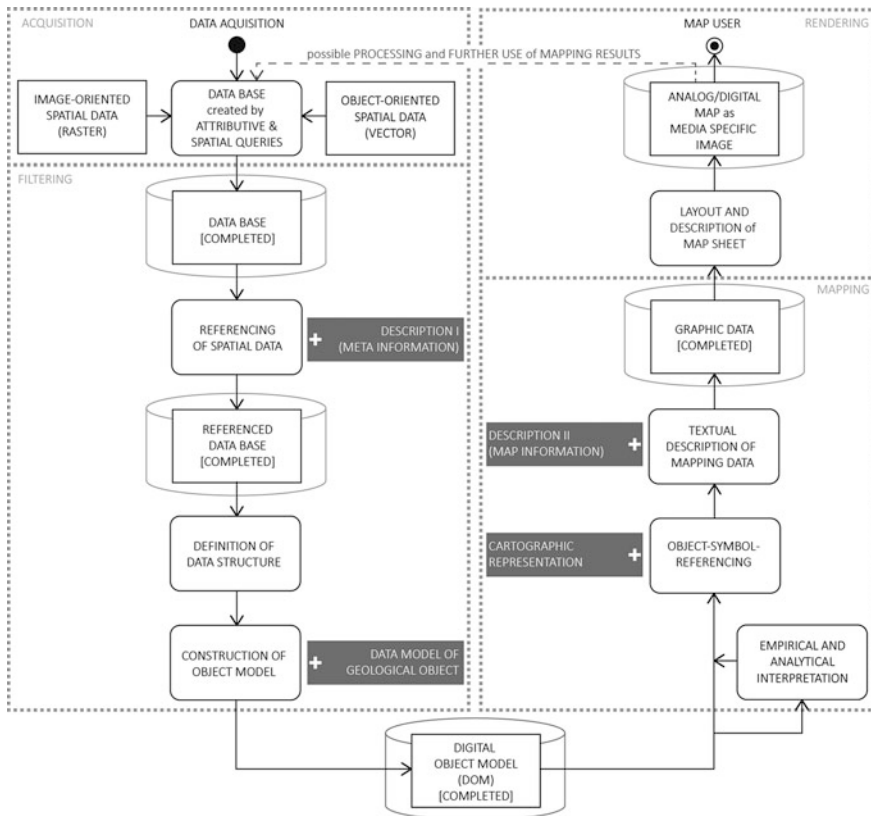
Fig. 1 GIS-based mapping process with integrated modules for an optimized workflow (Figure includes processinng steps defined by Haber 1990; Wood 1996; Carpeendale 2003)

- Map data must be *stored* in a well-structured way within any map data model. Properties of mapped entities are assigned during mapping and interpretation process (van Gasselt and Nass 2011a, b).
- Map data must be *visualized* through a homogenous and unambiguous object-symbol-reference. To achieve this, an adapted GIS-integrated symbol library is required which allows mappers to assign standardized sets of symbols for a homogenous appearance and exchange of map entities. This object-symbol reference must also be linked to the core data model in order to provide sustained functionality (Nass et al. 2011a, b).
- Map data must be *described* in order to trace back and review interpretation results. Such a description consists of entries comparable to a classic map legend that allows characterizing map layers. Another set of metadata descriptors is composed of information about the primary data basis (sensor data and auxiliary information) as well as the map product itself, including title, scale, mapping period, keywords, context (Nass et al. 2010).

In order to combine these three basic requirements within a common map-ping framework the basic model design should be independent of GIS architecture and implementation specification. As most COTS-and FOS-GIS rely on relational database management systems (RDBMS) a relational database model (DBM) is used to model dataflow. For database exchange in terms of DB structure as well as contents the XML interchange format (XMI) is used (OMG 2007). Signatures and symbols can be transferred to SVG (W3C 2003) and stored as accessible strings. For metadata description, XML (W3C 2008) provides a platform-independent description. The hierarchical structure of these formats can conveniently be decomposed into relations by means of XML-shredding (e.g., Freire 2003).

3.2 Mapping Guidelines

Thematic geologic or geomorphologic mapping, maps are usually not designed and compiled by cartographers but by researchers with a core expertise in a different field, such as geology or physical geography.

In order to obtain a decent level of cartographic fidelity, including map layout and basic design issues that allow proper communication and comparison, guidelines are needed which can be used for orientation and for cross-discipline references. For earth-oriented mapping, the "*Handbook on GIS and digital mapping*" provides a well-established compilation (United Nations 2000). For planetary mapping purposes baseline concepts have been formulated in the context of established systematic mapping programs (Tanaka et al. 1994, 2010). The focus of such guidelines is put on compliance issues with respect to organization a land administrative aspects rather than exact details of mapping conduct. Although several aspects concerning formatting are taken up, these are mainly focused on the overall map layout and map rendering rather than aspects of filtering and

mapping, i.e. delineation and assignments of symbols. In order to complement such work and in order to support the targeted cartographic representation of mapped entities it is necessary to provide basic rules and guides concerned with planetary GIS-based mapping, alignment and selection of symbols, and configuration as well as arrangement of map entities.

The following list comprises of few examples of such rules which are commonly violated if no framework has been established and communicated:

1. *Map Projections*: Georeferencing as well as assignment of coordinate systems must be fit to the map content and the methods that are to be employed during mapping and analyses. Issues with respect to the incorrect selection of coordinate systems are commonly encountered and such problems are difficult to resolve afterwards, in particular if data analyses have been carried out using map projections that are not capable of dealing with such analyses. On-the-fly projection methods do not suffice to maintain data fidelity.
2. *Map Scales*: Definition of a minimum map scale or map scale margins for mapping are required to prevent mapping at map scales that are not appropriate for depicting objects in a proper way on the final map sheet. Such issues require post-editing and time-consuming generalization. If required, it is recommended to provide space for inset maps conducted on larger map scales.
3. *Object Alignments*: Alignments of surface features and entities by means of representation layers must be organized in the way they occur in reality. It is important that the visibility and organization of objects is not disturbed by incorrect alignments of point symbols and area fill patterns (topologically different dimensions). Some of these issues can be avoided through topological rule sets or representations but the mapper must be aware of limitations despite technical support provided by the software system.
4. *Scale Bar*: Scale bar and numerical scales are required to allow map-based, i.e. analogue measurements.
5. *Legend*: In order to communicate and understand items depicted in the map, keys and legends are required which can be complemented by additional meta-information and text.
6. *Font symbols*: A map description helps to explain the map and map content directly. Placement, alignment and font styles/sizes have been well-defined within cartographic research. For non-cartographers, representation rules and guides must be employed to avoid common problems that cause misunderstandings (e.g., Hake 2002).

These basic rules are standard knowledge for cartographers; however, it has been experienced that these issues pose serious problems for non-cartographers and they are the main factors in project delay and data in-consistencies and integrity. These guides must be appropriately formulated within a mapping guide and it must be communicated how this mapping integrity can be reached by the technical means available in standard GIS.

4 Conclusion

The wealth of new planetary data arriving at research institutes and in agency archives requires means for handling and coordinating analyses and mapping. These tasks will occupy generations of researchers and cartographers as the amount of data as increased exponentially within the last 50 years. This increase puts on demands for new approaches in integrated data mapping which are only partially solved by making use of modern COTSGIS technology. The problem arises when cross-discipline communication and distributed collaboration are to be established. While the basic GI framework and data model can be extended in respond to user and organizational demands, the communication of demands from the side of cartographers to researchers and mappers is more complicated as no standards exist. Some cartographic issues must and can be organized through an appropriate data model; other issues must be addressed as guidelines for cartographic design.

References

Batson R, Whitaker E, Wilhelms S (1990) History of planetary cartography. In: Greeley R, Batson R (eds) Planetary mapping. Cambridge planetary science series, vol 6, Cambridge University Press, New York, p 12–59

Bleamaster L, Crown D (2010) Geologic map of MTM -40277, -45277, -40272, and -45272 quadrangles, eastern Hellas Planitiaregion of Mars. U.S. Geological Survey Scientific Investigations Map 3096, scale1:15,000,000, p 11

Buchroithner M (1999) Mars map—the first of the series of multilingual relief maps of terrestrial planets and their moons. 19thICA/ACI, Ottawa, Canada, p 1–3

Carpendale MST (2003) Considering visual variables as a basis for information visualization. Research report 2001-693-16, Department of Computer science, University of Calgary, Calgary

Chapman M (1999) Geologic/Geomorphic map of the Galindo quadrangle (V–40), Venus. U.S Geological Survey Scientific Investigations Map 2613, scale 1:5,000,000, p 13

Freire J, Siméon J (2003) Adaptive XML shredding: architecture implementation and challenges. University of Toronto, Toronto

Haber R, McNabb D (1990) Visualization idioms: a conceptual model for scientific visualization systems. In: Shriver B, Neilson GM, Rosenblum L (eds) Visualization in scientific computing. IEEE Computer Society Press, New York, p 74–93

Hake G, Grünreich D, Meng L (2002) Kartographie—visualisierung raum- zeitlicher informationen, 8th edn. Walter deGruyter, Berlin

Hargitai H (2006) Planetary maps: visualization and nomenclature. Cartographica 41(2):149–167. doi:10.3138/9862-21JU-4021-72M3

IAU (2012) International astronomical union. http://www.iau.org/. Accessed 8 Oct 2012

ICA (2003) A strategic plan for the International Cartographic Association 2003–2011. Adopted by the ICA Generally Assembly, Durban South Africa

Jaumann R et al (2007) The high resolution stereo camera (HRSC) experiment on mars express: instrument aspects and experiment conduct from interplanetary cruise through the nominal mission. Planet Space Sci 55(7–8):928–952. doi:10.1016/j.pss.2006.12.003

Lehmann H et al (2006) A thematic map of the centauri and hellas montes area, mars. 5th Turkish-German Joint Geodetic Day, Berlin

Nass A, van Gasselt S, Jaumann R (2010) Map description and management by spatial metadata: requirements for digital map legend for planetary geological and geomorphological mapping. In: 18th international research symposium on computer-based cartography. AutoCarto 2010, Orlando

Nass A et al (2011a) Implementation of cartographic symbols for planetary mapping in geographic information systems. Planet Space Sci 59(11–12):1255–1264, Elsevier Ltd., doi:10.1016/j.pss.2010.08.022

Nass A et al (2011b) Requirements for planetary symbology in geographic information systems. In: Ruas A (ed) Advances in cartography and giscience: selection from ICC 2011 lecture notes in geoinformation and cartography, vol 2. Springer, Berlin, p 251–266

OMG (2007) MOF 2.0/XMI Mapping, Version 2.1.1. OMG Doc Number: for- mal/2007-12-01. http://www.omg.org/spec/XMI/2.1/PDF

Neukum G, Jaumann R, HRSC-Team (2004) The high resolution stereo camera of Mars Express. ESA Special Publication, SP-1240, p 1–19

Neukum G et al (2009) HRSC: high resolution stereo camera. ESA Special Publication, SP-1291, p 15–74

Planetologia (2012) Planetary cartography. ELTE University Budapest. http://planetologia.elte.hu/. Accessed 16 Oct 2012

Roatsch T et al (2006) Mapping of the icy saturnian satellites: first results from cassini-ISS. Planet Space Sci 54(12):1137–1145. doi:10.1016/j.pss.2006.05.032

Roatsch T et al (2012) High resolution vesta high altitude mapping orbit (HAMO) atlas derived from Dawn framing camera images. Planet Space Sci 73(1):283–286

Seidelmann P et al (2005) Report of the IAU/IAG working group on carto- graphic coordinates and rotational elements: 2003. Celest Mech Dyn Astron 91:83–111

Shingareva K, Krasnopevtseva B (2001) Venus map—the series of multilingual maps for terrestrial planets and their moons. In: 20th ICA/ACI, Beijing, China, 5:3279–3284

Shingareva K, Krasnpevtseva B, Buchroithner M (2002) Moon map—a new map out of the series of multilingual relief maps of terrestrial planets and their moons. In: Conference GIS for sustainable development of territories. Petersburg, Russia, p 392–395

Shingareva K et al (2003) Mercury map—a new map out of the series of multi- lingual relief maps of terrestrial planets and their moon. In: 21th ICA/ACI, Durban, South Africa, p 1551–1554

Tanaka K (1994) The venus geologic mappers' handbook, 2nd edn. Open– File Report 94–438, Prepared for the National Aeronautics and Space Administration

Tanaka K, Skinner Jr J, Hare T (2010) Planetary geologic mapping handbook – 2010. U.S. Geological Survey, Astrogeology Science Center, Flagstaff, Arizona

United Nations (2000) Handbook on geographic information systems and digital mapping. United Nations publication, Studies in Methods, Series FNo. 79, ST/ESA/STAT/SER.F/79, New York

USGS (2012) Astrogeology science center. http://astrogeology.usgs.gov/PlanetaryMapping/. Accessed 11 Oct 2012

van Gasselt S, Nass A (2011a) Planetary mapping – The data model's perspective and GIS framework. Planet Space Sci 59(11–12): 1231–1242, Elsevier Ltd., doi:10.1016/j.pss.-2010.09.012

van Gasselt S, Nass A (2011b) Planetary map data model for geologic mapping. Cartography Geogr Inf Sci(CaGIS) 38(2):201–212. doi:10.1559/-15230406382201

W3C (2003) Scalable Vector Graphic (SVG) 1.1 specification. W3CRecommendation. http://www.w3.org/TR/2003/REC-SVG11-20030114/

W3C (2008) Extensible markup language (xml) 1.0, 5th edn. W3C Recommendation. http://www.w3.org/TR/2008/PER-xml-20080205

Wilhelms D(1990) Geologic mapping. In: Greeley R, Batson R (eds) Planetary mapping. Cambridge planetary science series, vol 6. Cambridge University Press, Cambridge, p 208–260

Wilhelms D, McCauley J (1971) Geologic map of the near side of the Moon. USGS Map I–703, 1:5,000,000, p 26

Wilhelms D, McCauley J, Trask N (1987) The geologic history of the Moon. U.S. Geological Survey Professional Paper 1348, United States Government Printing Office, Washington

Williams D et al (2011) Geologic map of Io. U.S. Geological Survey Scientific Investigations Map 3168, scale 1:15,000,000, p 25

Wood J, Brodlie K, Wright H (1996) Visualization over the world wide web and its application to environmental data. In: Proceedings of the 7th IEEE visualization conference, San Francisco, USA. doi:10.1109/VISUAL.1996.567610

On the Concept and Integration of Geologic Time in Planetary Mapping

Stephan van Gasselt and Andrea Nass

Abstract Planetary image data and maps form one of the most accessible scientific products for establishing cross-communication between planetary research disciplines and the general public. In particular geologic maps comprise a wealth of thematic information and form an accessible and esthetic medium for both, laypersons as well as scientist. Geologic maps form a substantial part of the planetary map data record that is publicly available. If such maps have been designed carefully they condense 2.5 + t dimensions (coll. 4D) into a two-dimensional map domain by connecting thematic attributes with geometry and time and by allowing (a) to completely reconstruct the subsurface extent as well as attitudes of mapped units by means of geometry, and (b) to establish a sequence of time units by relating legend items with geometric reconstructions. Despite the well-considered design of such maps, their higher non-geometric dimensionality and compression to two dimensions cause severe limitations in querying mixed non-spatial and spatial relationships, e.g., time, even in digital systems. This, however, is required for geological mapping in order to establish cross-relationships across regions on a local and even planetary scale. We here present a data framework which allows storing, managing and querying 2.5D + t information used in planetary geologic mapping. The focus is put on the general abstract ontological as well as the logical relationship concept which is designed to be employed in state-of-the art geographic information systems (GIS) commonly used for planetary geologic mapping.

Keywords Planetary mapping · Geology · Cartography · Time-integration · Data models · GIS

S. van Gasselt (✉)
Institute of Geological Sciences, Freie Universitaet Berlin, Berlin, Germany
e-mail: vanGasselt@gmail.comStephan.vanGasselt@fu-berlin.de

A. Nass
Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany

1 Introduction and Motivation

Planetary geologic maps are cartographically sophisticated products depicting a four-dimensional geologic content on a two-dimensional sheet of paper. This was the case in 1815 when first geological maps of Great Britain were created by Smith (1815), and this remains exactly the same 200 years later when geologic maps covering our planetary neighborhood at different map scales have become widely available.

Due to their inherent geologic scientific information, such special-purpose maps are well-established means for compressing geometrically as well as thematically higher-dimensional contents onto a flat print-out paper for easy access, convenient storage and distribution. In these respects, planetary maps are not different from earth-oriented maps apart from the fact that earth-oriented maps are usually validated—at least locally—due to focused in situ field investigation while planetary maps are solely based upon remote-sensing image-data interpretation. Until today, the geologic map record stored in the digital and analog archives of the United States Geological Survey (USGS) amount to several hundred published maps of Mercury (e.g., Grolier and Boyce 1984), Venus (e.g., Chapman 1999), the Moon (e.g., Wilhelms and McCauley 1971), Mars (e.g., Bleamaster et al. 2010), and the Galilean Satellites (e.g., Williams et al. 2011). Additionally, some of the Soviet-time geologic maps (beside other map products) have recently become available in digital formats and it is anticipated that the Russian archives at MIIGAiK and at other places host a plethora of thematic geologic maps (Hargitai, 2006; Hargitai et al. 2010).

The significance of geologic maps as scientific tools in planetary exploration remains unchallenged and during the last decade long-established technical means in the field of Earth-oriented GIS-based mapping have finally entered the stage of systematic planetary geologic mapping. Despite technical advances, GIS technology has predominantly been used for providing convenient means for digitizing, integrating raster and vector based datatypes and for data exchange while the possibilities with respect to the data-management backend remained untouched.

For geology, surface materials and their stratigraphic name codes and time relationships play a fundamental role in order to establish a spatio-temporal understanding that is inherently related to the field of geology. Consequently, these spatio-temporal aspects are required to be reflected in the proper assignments of attributes within any GIS-based data model that allows being administrated, queried and—finally—visualised as a geologic map. Once established such data models allow users to efficiently query attributes and establish a multidimensional approach for knowledge extraction covering time, space and attribute dimensions. The extraction of data and generation of knowledge can be targeted to large scale, i.e. localized, areas as well as small scales, i.e. global relationships and—if consistently designed—it allows building queries across planets while maintaining a single data model.

The aim of this work is to establish such a time-integrative planetary geologic data model which is specifically targeted at the design of time-relationships. This is achieved by building an abstract understanding of geologic time concepts (see chapter "Implementation of Cartographic and Digital Techniques in Orienteering Maps") and by integrating these concepts into a cartographic framework for planetary geologic mapping (see chapter "Map Projection Reconstruction of a Map by Mercator"). Few example queries are demonstrated using a map example (see chapter "The Pole is Impracticable but *There* is a Land Northward: Austro-Hungarian Pole Expedition and Mapping of the Franz Josef Land"). This work complements a proposed planetary-mapping data model with respect to full time integration (van Gasselt and Nass 2011a, b).

2 Time and Space in Planetary Geologic Mapping

Planetary geologic units are rock-materials that have been formed by a discrete process in a discrete time interval (Wilhelms 1990, p. 210). Geologic maps consequently depict the spatial extent of geologic materials, and if the relative vertical position of material units is known the relative temporal relationships can be established as stated by the law of superposition (Steno 1916). The three-dimensional extent of a given material thus defines the geologic unit while the interrelationships of different units define the timeframe. If such observations are carefully conditioned, transferred to a proper map layout and complemented by topographic contour lines, attitudes and sub-surface extent of geological units can be geometrically reconstructed from the map sheet.

The importance of geologic maps lies in the concept that the arrangement of geologic units depicts a sequence of events, generally summarized as stratigraphy. If the sequence is based upon the relative arrangement of rock-units with older units being buried by younger ones, the concept is termed *rock-stratigraphy* or *litho-stratigraphy* (Fig. 1). If the time, units have been formed and deposited is known, a *time-stratigraphy*, or *chrono-stratigraphy* can be established (Wilhelms et al. 1990; Greeley and Batson 1990). The concept is relatively straightforward, the problem, however, arises as soon as an actual time stamp for a mapped unit has to be assigned.

Due to limited accessibility of planetary environments, the absolute measurement of an age for a given unit is problematic and apart from radiometric ages for meteorites and returned lunar rocks in the course of the Apollo and Luna program, a detailed assessment of ages is missing. This limitation was overcome when the unambiguous identification of an abundance of impact craters on planetary surfaces led to the idea that there frequency can be used to assess the time that a unit has been exposed to the asteroidal bombardment. The more impact craters of a given diameter size are identified, the older a unit is. If material units are properly delineated and impact-crater size-frequencies for each unit are known, again, a relative timing can be established. For absolute age assignments, however, each
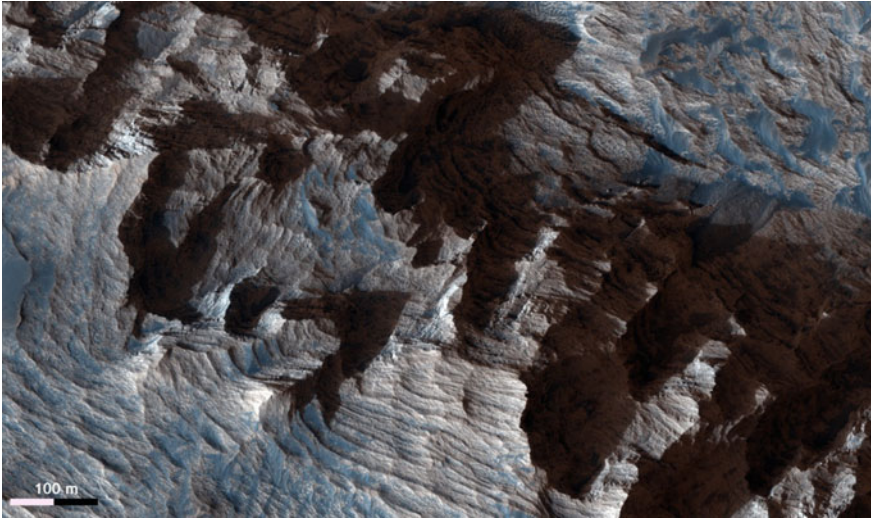
Fig. 1 Geological layering in Candor Chasma, Mars, as imaged by the High Resolution Science Experiment HiRISE onboard Mars Reconnaissance Orbiter. The vertical stack of layers depicts the relative litho-stratigraphy. Scene width is approximately 1 km across, North is up. Press Release PIA13725, 2007, NASA/JPL Caltech/University of Arizona

planet's impact-crater size-frequency record must be transferred to absolute ages. This could be accomplished by comparing remotely-sensed impact crater populations with radiometric ages derived from sample data for the Moon to establish a chronology. For all other planetary objects, the lunar chronologies had to be transferred under consideration of the respective space environment, impactors and target properties (Holsapple 1987; Schmidt and Housen 1987). Currently, a formal stratigraphic scheme has been developed for each of the inner planets (Tanaka and Hartmann 2008). Amore detailed treatment and a summary of planetary chronologies and aspects of chronostratigraphy along with literature references are provided in Ogg et al. (2008), van Gasselt and Neukum (2010).

The majority of scenarios in planetary geology and geomorphology can be associated with time as time is associated with each spatial entity, either *implicitly* or *explicitly*, due to the fact that geology *is* constrained and defined by time (see, e.g., Gould 1987). Time in geography and congeneric disciplines can be considered in different ways as it is either a linear continuum without a distinct beginning and an end or it may be represented by discrete steps with pre-defined time-resolution intervals (see also Pred, 1977; Langran, 1992; Peuquet, 1994; Peuquet, 1999; Raper, 2005; Le and Usery, 2009).

A cartographic data model must cope with queries related to the geometry of entities as well as time information and relies on an approach that incorporates both, spatial and time geometry equivalently. Time is often represented as an irreversible continuum with a starting point indicating, e.g., planet formation and

the presence as terminal point. This concept applies to all geologic time scales for all terrestrial solar-system bodies, i.e. planets, moons and asteroids. Thus, geologic time is linear and absolute in nature (Le and Usery, 2009; Claramunt and Theriaupt, 1995) and is contrasting to the concept of cyclic or periodic time known from, e.g., recurring changes of seasons. Planetary geologic time, i.e. the chronology, as internationally defined, is partitioned into successively smaller levels of time units with distinct age boundaries. This time is absolute and thus measurable as chronometric time before present (see also Salvador, 1994; Gradstein, 2004; Couvering, 2007; Ogg, 2008). This leads to another distinct feature of geologic time: time is measured backwards as geologic events and changes (boundaries) occurred before present time. Consequently, geologic boundaries are dynamic with respect to today and static with respect to a pre-defined reference date.

Information put on a cartographic product such as a geologic map targets at depicting two issues at the same time: the spatial extent of materials and rock units exposed at the surface and the temporal extent with respect to the chronostratigraphic context by relating spatial units to an age. Usually, this is done by colour-coding and an associated map legend. Without a legend, stratigraphic relationships, i.e. the sequence of deposited or emplaced material can only be reconstructed from the spatial relationships of map units and their topographic position.

Thus, the need for time integration lies in the nature of geologic mapping and time attributes establish a link for answering questions such as

- The time one or more units were emplaced (absolute or relative time).
- The correlation of time-units (chronologies) across planetary bodies (absolute time).
- The names (and attributes) of geologic formations deposited at a time in a given region (absolute time).
- The names (and attributes) of geologic formations situated below or above another formation (relative time).

In order to work correctly such data models do not only require time associated with the geologic contents but also with respect to the moment a particular content has been valid (validity time) as well as with respect to the time, the data contents were created (database time).

3 An Explicit Time-Integrative Approach

A time-integrative data model has been conceptualized and physically implemented using ESRI's commercial software suite ArcGIS using a file-based geodatabase (FGDB) as well as server environment using the ArcSDE connector. The overall design (Fig. 2) is evolved around the geometry represented as feature class *Geological Unit* (GU) which inherits positional IDs to facilitate extraction of *on-top-of* or *below-of* queries, and the lateral geometry. By using recursive relationships the rock-stratigraphy, i.e. the relative sequence of units, is extracted.
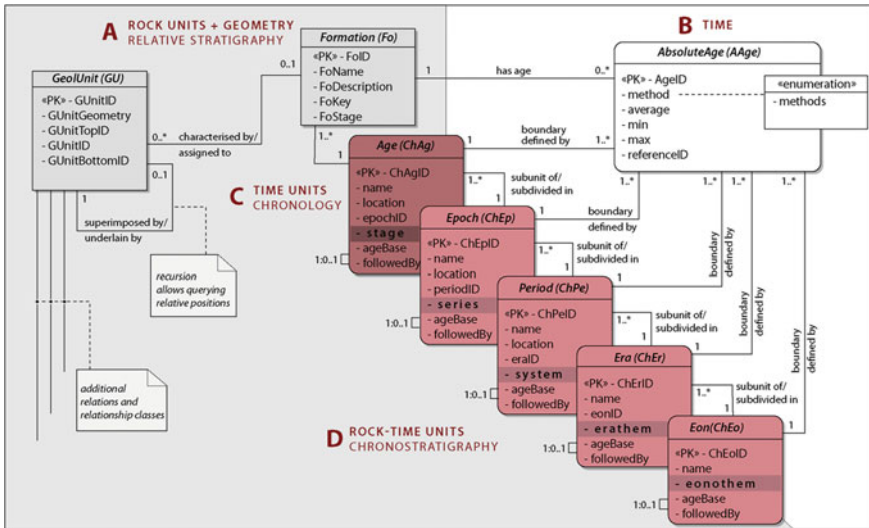
Fig. 2 Entity Relationship Model (Data Model) connecting stratigraphic units (relative scale, *A*) with absolute ages (*B*) and the chronology and chronostratigraphy (*C* + *D*). The model core is evolved around the geologic unit and its geometry (GU)

Attributes related to time, i.e. chronology and chronostratigraphy, are separated as these are prone to change. The geologic formation (*Fo*) as core entity has one or more absolute ages as measured by one or more age-determination methods (*AAge*). These methods can be based upon impact-crater size-frequencies and a chronology model, or they can be based on radiometric age determinations. Each of the methods has an inherent accuracy defined through minimum and maximum ages that allow querying age ranges. The hierarchically organized chronostratigraphy is built using a relational cascade, starting with the smallest entity (*ChAg*) representing the lowest age level. For space reason we here demonstrate two small example settings (Fig. 3) that involve (1) a time query which returns a stratigraphic table, and (2) a spatial query using time conditions returning object ID sand selections within a GIS.

A map of the lunar Mare Orientale basin was constructed and synthesized using previous analogue mappings (mainly map sheet I-1034). The actual extent of features is less important than the actual attributes and implicated stratigraphy. The map is therefore accompanied by a stratigraphic column (Fig. 3).

Most of the geologic units belong to the Imbrian (lithostratigraphic) System (Orientale Group), denoted by I. The Orientale basin material (I) is superimposed by younger impact-crater materials from the Imbrian (blue, Ic_2), Eratosthenian mare material (Elm), Eratosthenian impact crater materials (green, Ec) and impact material from the Copernican Systems (yellow, Cc). Without going into the details of the Orientale-Group stratigraphy (all units are considered to be contemporary), three generations of impact craters and Mare material units can be traced (Fig. 3).
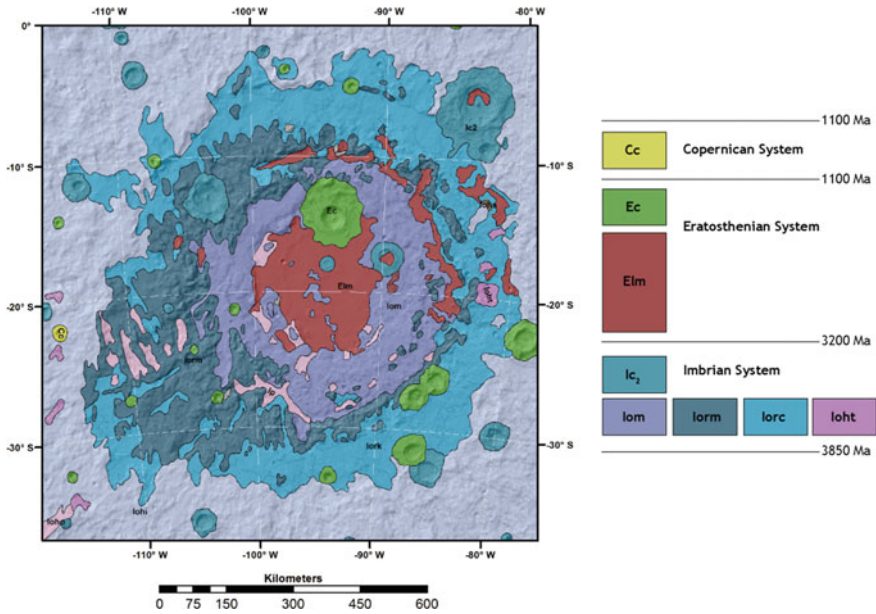
Fig. 3 Geologic map of the lunar Mare Orientale basin as an example of a planetary geologic map that has originally been crafted for analogue distribution in 1970 based on Apollo and Lunar Orbiter data (USGS). Shaded relief background is based upon Lunar Reconnaissance Orbiter data. Figure shows the geologic map and associated high-level stratigraphy (systems/eras only) for demonstration of example queries

The stratigraphic systems are transferred to chronostratigraphic eras using the recent chronology model as described in Tanaka & Hartmann (2008). For demonstration purpose, lower-level stratigraphic subdivisions (series) and chronostratigraphic subdivisions (periods) have been omitted in the map example. For a complete treatment of the geology of Mare Orientale, see Wilhelms, 1987, pp. 57–82.

Case 1. Hierarchical Stratigraphic Relationships

Chronostratigraphic divisions and subdivisions are obtained using a cascading rightouter join query. The given example descends from the stratigraphic level (here, the Eratosthenian) downwards and can be easily extended using higher-level subdivisions if available. Such a query resembles the extraction of a stratigraphic chart for a given chronology model.

```
SELECTChEp.Name AS 'Epoch',
ChPe.Name AS 'Period',
ChEr.Name AS 'Era',
FROM ChEp
RIGHT OUTER JOIN ChPe ON ChEp.periodID=ChPeID
RIGHT OUTER JOIN ChEr ON ChPe.eraID=ChErID
RIGHT OUTER JOIN ChEo ON ChEr.eonID=ChEoID
WHERE ChEr.Name="Eratosthenian"
ORDER BY    ChEr.AgeBase DESC,
ChPe.AgeBase DESC,
ChEp.AgeBase
```

The return is a table consisting of three stratigraphic columns (epoch—period—era) with their sorted contents and, if required, associated ages. If formations are added, the names of formations are returned. In order to be conveniently applicable for end users, such a query should be built from a predefined query stack. Variables (era and the level of hierarchy) must be parsed from the user interface.

Case 2. Identification of Units based on Ages

More important than the extraction of ages and age cascades is the identification of spatial units associated with an age. An example query could ask for all units of a given age or within a given age range. Alternatively, spatial objects can be identified using *on-top-of* or *below-of* relationships as defined through a recursive query on GU.

If we assume we want to identify on the map all units that are older than, e.g., 3200 Ma, i.e. pre-Eratosthenian, or that are younger than materials in the stratigraphic unit Ec (which results in a selection of Copernican-System impact craters) an example query reads as follows.

```
SELECT GUnitIDAS 'ID',
FROMGU
  RIGHT OUTER JOIN FoONFo.FoID=GU.GUnitID
    RIGHT OUTER JOIN ChAgON Fo.FoStage=ChAg.Stage
RIGHT OUTER JOIN ChEp ON ChAg.epochID=ChEp.ChEpID
        RIGHT OUTER JOIN ChPe ON ChEp.periodID=ChPe.ChPeID
. . .
            WHERE ageBase>3200000000
              ORDER BY  GUnitIDDESC
```

As a result the ID of each spatial unit (GU) is returned that is true with respect to the query, i.e. all pre-Eratosthenian feature classes. This ID is then parsed to the selection within the GIS environment for subsequent analysis. If the age-dating method requires to be incorporated, an additional right outer join on *AAge* is inserted. Again, such a task can only be efficiently implemented if a predefined query skeleton is provided. Variables parsed as user input are: age value and

operator ($<$, $>$, $\geq$, $\leq$). Appropriate user interfaces for convenient (and more intuitive) time queries are currently being developed and implemented.

Once the framework is set up, it is a matter of filling relations with data and constructing queries that relate data to each other in an appropriate way. These are just two simple examples of querying attributes (e.g., for reports) and IDs in order to extract geometries (feature classes) from the geodatabase. The feature class itself only stores geometry, self-referencing feature-class IDs and an ID pointing towards the formation relation. All other relations are considered auxiliaries.

4 Summary and Prospects

A geologic mapping time model should be as generic as possible and as detailed as necessary in order to pose no limitations for incorporation of future requirements and changes. In particular it should be applicable to all other planetary objects for which geologic and stratigraphy concepts as well as hierarchies and vocabulary are equivalently organized. The knowledge about Earth's stratigraphy is much more detailed and the chronostratigraphic and stratigraphic hierarchy descends much further down to the level of few decades and few centimeters of layer thicknesses, respectively. In particular age determinations differ significantly for Earth cases as radiometric ages have been collected for many geologically significant areas and not only for selected sample collections. The data model, however, allows to be expanded towards a higher granularity in order to cope with terrestrial demands.

While the query-selection interaction shows currently no known limitation with respect to relations and data, the usability requires to be significantly improved to provide an intuitive access for queries that are not expressed in a dedicated query language. Furthermore, database setup is being developed using automated procedures but the possibilities for joining and querying relations using a file-geodatabase (FGDB) concept are limited so that a server option seems to be the best alternative. This would also allow transferring the model for other (possibly open-source) GIS flavours.

References

Bleamaster LF, Crown DA (2010) Geologic map of MTM-40277, -45277, -40272, and -45272 quadrangles, eastern Hellas Planitia region of Mars: U.S. Geological Survey Scientific Investigations Map 3096, 11 p, scale 1:15,000,000

Chapman MG (1999) Geologic/geomorphic map of the Galindo Quadrangle (V–40), Venus: U.S. Geological Survey Scientific Investigations Map 2613, 13 p, scale 1:5,000,000

Claramunt C, Theriault M (1995) Managing time in GIS: an event-oriented approach. In: Clifford J, Tuzhilin A (eds) Recent advances in temporal databases: proceedings of the international workshop on temporal databases. Springer, Zürich, pp 23–42

Couvering JA, Ogg JG (2007) The future of the past: geological time in the digital age. Stratigraphy 4:253–257

Gould SJ (1987) Time's arrow and time's cycle, Myth and metaphor in the discovery of geological time

Gradstein FM, Ogg JG, Smith AG et al (2004) A geologic time scale. Cambridge University Press, Cambridge

Greeley R, Batson RM (1990) Planetary mapping, Cambridge planetary science series 6. Cambridge University Press, Cambridge

Grolier MJ, Boyce JM (1984). Geologic map of the Borealis region (H-1) of Mercury, USGS Miscellaneous investigations series Map I–1660, 21 p, 1:5,000,000

Hargitai HI (2006) Planetary maps: visualization and nomenclature. Cartographica 41(2):149–167. doi:10.3138/9862-21JU-4021-72M3

Hargitai HI, Shingareva KB, Golodnikova IY, Gede M (2010) Historic soviet planetary maps online'. In:' 41st lunar and planetary science conference abstracts, LPI contributions, #1209

Holsapple KA (1987) Impact crater scaling laws. Intl J Impact Eng 5:343

Langran G (1992) Time in geographic information systems, ser. technical issues in geographic information systems. Taylor & Francis, London

Le Y, Usery EL (2009) Adding time to GIS. In: Madden M (ed) Manual of geographic information systems. The American Society for Photogrammetry and Remote Sensing, Bethesda, pp 311–332

Ogg JG, Ogg G, Gradstein FM (2008) The concise geologic time scale. Cambridge University Press, Cambridge

Peuquet DJ (1994) It's about time—a conceptual framework for the representation of temporal dynamics in geographic information systems. Ann Assoc Am Geogr 84:441–461

Peuquet DJ (1999) Time in GIS and geographical databases. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds) Geographical information systems, 2nd edn. Wiley, Chichester, pp 91–103

Pred A (1977) The choreography of existence: comments on Hägerstrand's time geography and its usefulness. Econ Geogr 53:207–221

Raper J (2005) Multidimensional geographic information science. Taylor & Francis, London

Salvador E (1994) International stratigraphic guide, 2nd edn. Geological Society of America and the International Union for Geological Sciences (IUGS)

Schmidt RM, Housen KR (1987) Some recent advances in the scaling of impact and explosion cratering. Int J Impact Eng 5:543–560

Smith W (1815) A delineation of the strata of England and Wales with part of Scotland; exhibiting the collieries and mines, the marshes and fen lands originally overflowed by the sea, and the varieties of soil according to the variations in the Substrata. London: J Cary, scale ca. 1,313,800

Steno N (1916) The prodromus of Nicolaus Steno's dissertation concerning a solid body enclosed by process of nature within a solid; an English version with an introduction and explanatory notes by J. G. Winter, Univ. Michigan Humanist. Ser., XI, New York, Macmillan (1669)

Tanaka KL, Hartmann WK (2008) Planetary time scale. In: Ogg JG, Ogg G, Gradstein FM (eds.) The concise geologic time scale. Cambridge University Press, New York, pp 13–22

van Gasselt S, Nass A (2011a) Planetary mapping—the datamodel's perspective and GIS framework. Planet Space Sci 59(11–12):1231–1242

van Gasselt S, Nass A (2011b) Planetary map data model for geologic mapping. Cartogr Geogr Inf Sci 38(2):202–213

van Gasselt S, Neukum G (2010) Chronology, cratering and stratigraphy. 'In:' M. Gargaud, R. Amils, J. Cernicharo Quintanilla, H. J. Cleaves, W. M. Irvine, D. Pinti, M. Viso (eds) Encycloepdia of Astrobiology, Springer, Berlin

Wilhelms DE (1987) The geologic history of the Moon. U.S. Geological Survey, Washington, p 1348

Wilhelms DE (1990) Geologic mapping. In: Greeley R, Matson RM (eds) Planetary mapping, Cambridge planetary science series 6. Cambridge University Press, Cambridge, pp 208–260

Wilhelms DE, McCauley JF (1971). Geologic map of the near side of the Moon, USGS Map I–703, 26 p, 1:5,000,000

Williams DA, Keszthelyi LP, Crown DA et al (2011) Geologic map of Io: U.S. Geological Survey Scientific Investigations Map 3168, 25 p, scale 1:15,000,000

Part VI
Cartography and Environmental Modelling

Reservoir Water-Transparency Mapping by Means of Multispectral Ikonos Imagery

Adriana Castreghini de Freitas Pereira

Abstract In current society, drinkable water has been the subject of innumerable debates, mainly in scientific groups, in which, through researches focused on water availability and quality, it is possible to prepare diagnoses and point out solutions to planners and decision makers. The water-transparency, beyond than being a physical feature easily obtained on field, also represents a correlation with the superficial electromagnetic radiation from the water body, enabling its assessment both by multispectral images taken by sensors from orbiting platforms, as spectral data collected in situ. The purpose of this research was to realize the inference of water transparency, using a multispectral IKONOS imagery, at the bands 1 (450–520 nm); 2 (520–600 nm); 3 (630–690 nm); and 4 (760–900 nm) and spectral data collected in situ with FieldSpec UV/VNIR (400–900 nm) spectroradiometer. After initials data analysis and processing, the whole data was submitted to correlation analysis, developing an inference model of water-transparency. In conclusion, the specific objectives was achieved by the inference model of water-transparency through $\left(\frac{B4}{B1}\right)$ band ratio of Ikonos multispectral images, corresponding to 450–520 nm (B1) and 760–900 nm (B4). The spatial distribution of inference model was accomplished and the Secchi Depth results evinced consistency with those obtained in situ to central and west regions of the Reservoir. This indicates that the inference model of water-transparency was appropriated. This idea is reinforced on the statistical validation of the model.

Keywords Remote sensing of water · Water-transparency · Statistical inference

A. C. de Freitas Pereira (✉)
Department of Geosciences, Science Center, Londrina State University—UEL, Londrina, Brazil
e-mail: adrianacfp@uel.br

1 Introduction

In current society, drinkable water has been the subject of innumerable debates, mainly in scientific groups, in which, through researches focused on water availability and quality, it is possible to prepare diagnoses and point out solutions to planners and decision makers.

The water-transparency is one of the physical parameters observed on water bodies' quality analysis. This parameter can be achieved in the Secchi disc fathom, in other words, by observing the disappearance of a white disc stepped in water. This value has a direct relation with the water-transparency and inverse relation by the amount of organic and inorganic compounds in the light path and also with the attenuation coefficient of irradiance (Pereira Filho 2000). This means that the Secchi depth disappearance corresponds to those 400–700 nm (visivel range) depth of radiation reflected, which is no longer sensitive to the human eye (Esteves 1998).

The water-transparency, beyond being a physical feature easily obtained on field, also represents a correlation with the superficial electromagnetic radiation from the water body, enabling its assessment both by multispectral images taken by sensors from orbiting platforms, as spectral data collected in situ.

Generally, scientific researches related to water quality use, among other available methods on environmental analysis, the analysis of correlation between the optically active components concentration and the spectral information of water bodies from laboratories and obtained in field (Rundquist et al. 1996; Louchard et al. 2002; Nobrega 2002; Barbosa 2005; Rudorff 2006). Others authors emphasize the estimates of concentration from this components by orbital data (Dekker 1993; Ritchie and Cooper 1998; Giardino et al. 2001; Galvão et al. 2005; Novo et al. 2006; Ekercin 2007; Pereira et al. 2011).

In both cases, the empiric models development attempting to estimate some of the components by other components observed in field, allow spatial representation with reduced costs, reducing also the lab tests.

In this context, the aim objective of this research was to realize the inference of water-transparency by multispectral IKONOS imagery and spectral data collected in situ, developing the water-transparency mapping of Itupararanga Reservoir/SP/BR.

2 Approach and Methods

Aiming to develop the research, a high spatial resolution multispectral image from the IKONOS II Satellite (spectral bands 1, 2, 3 and 4, corresponding respectively to the intervals of wavelength between 450 and 520 nm; 520 and 600 nm; 630 and 690 nm; 760 and 900 nm) was obtained simultaneously to the performance from a field survey, in which was measured the parameter of Secchi Depth and the

spectral curves with FieldSpec spectroradiometer UV/VNIR (Analytical Spectral Devices, Inc.) on sample points previously defined.

Spectral data obtained by the spectroradiometer was made with 25° viewing angle covering an area of 0.22 m^2 and the Reflectance Factor was calculated measuring the radiance from a reference plate Spectral on with 99 % of constant reflectance in all the wavelength (400–900 nm) for lab calibration.

The spectral radiance from water in each point was obtained by the average of three integrations, trying to keep the same viewing geometry at all points.

The correlation analysis was realized between data collected in situ (Secchi Depth and spectral curves) and the multispectral image. Classic Statistic regression methods were applied in order to generate water transparence inference models, estimating the spatial distribution of water-transparency to models which consider the multispectral image.

Figure 1 exposes the research study area, which was Itupararanga's (SP/BR) Reservoir, belonging to Sorocaba's River Watershed, one of the largest springs of drinking water from Sorocaba's region, considered as good quality water by the
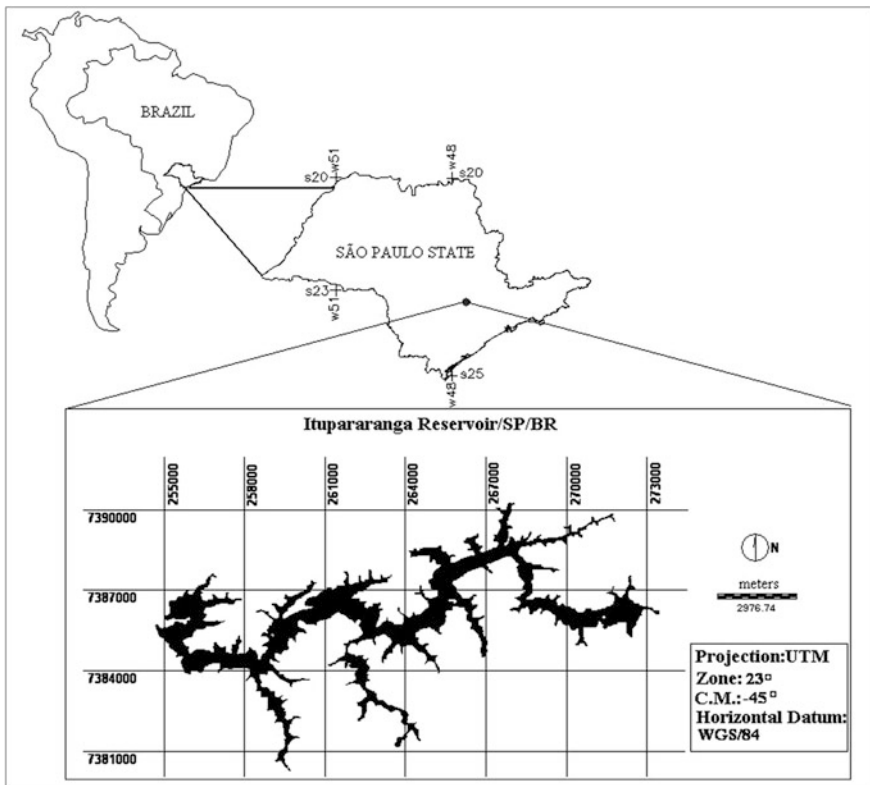
Fig. 1 Itupararanga reservoir localization in São Paulo State, Brazil

official bureaus, supplying the cities of Votorantim, Mairinque, Alumínio, Ibiúna and São Roque, in an approximate amount of 800.000 inhabitants.

2.1 General Aspects

Planning the field work is one of the most important steps to acquire the spec-troradiometry data in water bodies, because the incident sunlight and the water's surface characteristics control the light field submerged, and also the degree in which spectral data represent the optically properties from water bodies. In this context, the meteorological conditions have a directly influence on the optically properties of water bodies, and also on the quality of spectral information.

The water-transparency data collected in situ was based on a pre-defined sample scheme to the distribution of the sampling elements. This scheme con-sidered a number of spots that enabled the samples optimization and guaranteed, at the same time, spatial representativeness from the parameter collected for water quality analysis. Therefore, it was based on the water's body spectral variability during a period of time and by the entrance of the main affluents into the Reservoir (Pereira et al. 2007). All in all, 72 sampling sites were defined, spreaded throughout the Itupararanga's/SP/BR Reservoir, and also on the tributaries entrances, as shown by the Fig. 2.

The Secchi depth survey in situ took place on February 7th and 8th/2007, in such a wise as illumination conditions (nebulosity) and wind (waves on water
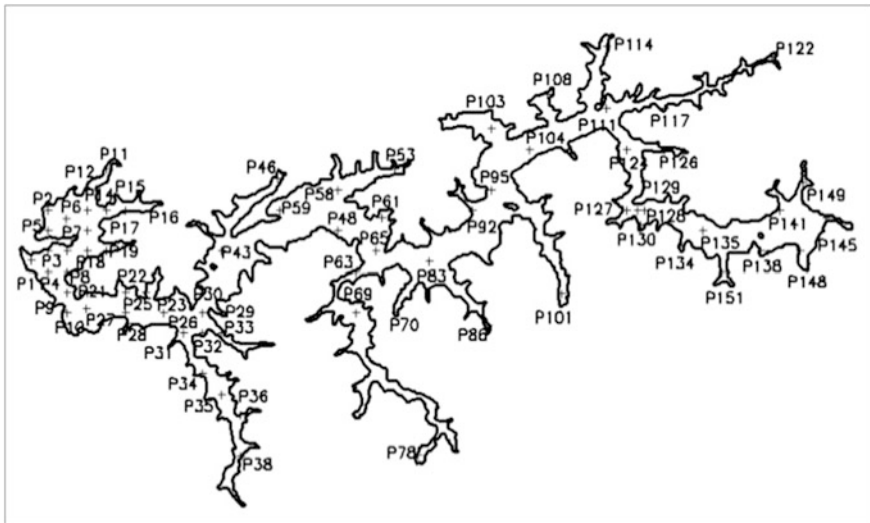
Fig. 2 Itupararanga reservoir/SP/BR and the localization of 72 points defined for a mostral scheme for data collected "in situ"

surface) were registered on field spreadsheets. At the first day on field, measurements were performed in 31 sampling elements, at sun, clouds and winds presence. By the second day, only four sampling elements were evaluated, on torrential rains and moments of drought with clouds occurrence, totalizing 35 sampling spots.

Simultaneously to the Secchi depth survey on field, a multispectral image of high spatial resolution from the study area was obtained by Ikonos II satellite, corresponding to the bands B1 (spectral range of 450–520 nm); B2 (spectral range of 520–600 nm); B3 (spectral range of 630–690 nm) and B4 (spectral range of 760–900 nm). These images can be programmed for the period wished by the user, however, it can also be taken without specific definition of the days. Besides, in due to the need of synchronism between the obtainment of Secchi depth parameter and the obtainment from the image, a definition of a better date to the collection can be just approximate. Therefore, the survey in situ and the obtainment of the image was planned for the week of February 5th to 9th, having the image taken by the Ikonos II Satellite in February 5th, thus, lagged in 2 or 3 days from the obtainment of the scene. The image obtained was considerably covered by clouds in most parts of the concerned spots from the research.

2.2 Collection of the Spectral Data and Secchi Depth in situ

Due to the fact that clouds partly covered the image the sampling sites and also, the heterogeneity of weather conditions in the two days of field works, there was a need to classify the sample in four groups, following meteorology conditions by the moment of its obtainment, as shown at Table 1.

Table 1 Classification of sample points according to meteorology conditions during the survey in situ and clouds in Ikonos imagery (author)

	Group 1	Group 2	Group 3	Group 4
Samples	03	10	12	10
Points	P30, P32, P36	P22, P23, P25, P26, P27, P28, P29, P33, P34, P35	P2, P5, P6, P7, P8, P9, P10,P12, P15, P16, P19, P31	P1, P3, P14, P17, P18, P21, P104, P111, P117, P129
Sky	Clear	Clear	Cloudy	Cloudy
Wind	Light	Average	Light	Strong
Wave	Wavelet	Wavelet/medium	Wavelet	Medium/high
Hour	14:51–15:32	14:03–16:06	11:32–15:12	10:58–14:10
Localization	Western sector	Western sector	Extreme western sector (dam)	Western and eastern sectors
Average Secchi depth (m)	2.20	2.37	2.45	2.13
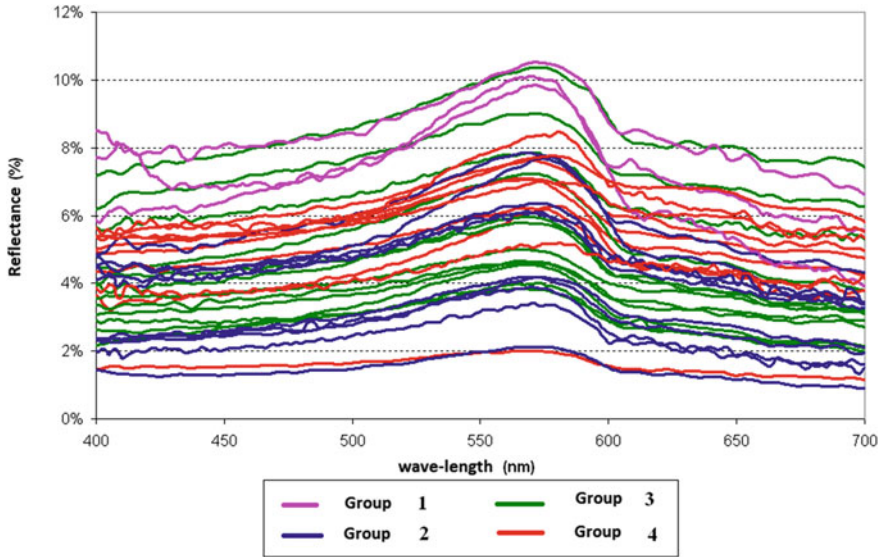
Fig. 3 Smoothing reflectance curves to each group

To the spectral curves from each sample element, obtained by field spectro-radiometer (Tsai and Philpot 1998; Rudorff 2006) a smoothing filtering was applied to reduce the effects of noises observed in wavelength interval from 400 to 700 nm. Mobile filters of smoothing were tested on average of three, four, five, six, seven and nine points besides the smoothing from Fourier.

The chosen smoothing filtering to each curve was the one which reduced the random noise without changing the position of dispersion and absorption features (λmax e λmin) and the module derived (maximum value and minimum value), according to Pereira (2008). The smoothing reflectance curves of points belonging to each group are presented in Fig. 3.

Figure 3 shows that the sample points from the four groups have similar spectral behavior, with ranging reflectance between 2 and 10 %. None of them presents significant reflectance on blue lightband (400–450 nm) probably due to the presence of wind, which reduces the light transmission in interface air/water. Although the geometry of acquisition from the data collected was oriented to reduce the water specular component, the surface sign was lesser than in plan surface water conditions, or conditions without wind.

A maximum reflectance peak (between 6 and 10 %) was observed at a range of green (560–570 nm) in many of the sample points, what could be associated to the presence of chlorophyll, according to Kirk (1994), Rundquist et al. (1996) and Barbosa (2005).

The techniques of spectral analysis constituted through band ratio and difference were applied in smoothing spectral curves in spectral intervals that better characterized the components of the Itupararanga water body researched in each of

the four groups. For this, the average of minimum and maximum reflectance was calculated, applying the first derivate in all smoothing spectral intervals that defined features of absorption and scattering, which were: 571–698 nm for the group 1; 571–696 nm for the group 2; 568–648 nm for the group 3 and 574–584 nm for the group 4. The ratio operations considered the reflectance in points: (571/698, 571/696, 568/648, 574/584) and the differences (571–698, 571–696, 568–648, 574–584), being applied to each of the four groups. The data were normalized as a logarithmic function.

3 High Resolution Multispectral Image

The multispectral image from Ikonos II, corresponding to bands B1, B2, B3 and B4, with respective spectral intervals of 450–520, 520–600, 630–690 and 760–900 nm, was submitted to treatments in order to make it spatially and radiometrically compatible with the Secchi Depth parameter obtained in situ. The treatments were such as: scenes georeferencing, clipping mask from the water body and radiometric and atmospheric corrections.

The image georeferencing was made by the program SPRING-INPE/BR (Georeferenced Information Processing System—National Institute for Space Research, Brazil), using the Affine Transformation on the scheme and 12 points of control. The residue of the geometric transformation was 1 pixel, considered acceptable for the work.

The delimitation of the Reservoir was made on SPRING, by choosing segmentation from region of growth using similarity parameters equal to 20 and area of 200 pixels, after tests made with other values, due to the fact that it is the most appropriate way to delimit the contours of the dam, comparing to the others. It was applied the technique of unsupervised classification Isoseg to group the defined regions in the segmentation with a threshold value equal to 90 %. The mapping from the thematic classes was performed, and also the manual edition of the contour. The developed mask was used to cutout the Information Plans regarding the four bands of the multispectral image. The radiometric calibration was performed with the objective of converting the DN's (digital numbers) from the original image into spectral radiance. For this, the spectral radiance or sensor radiance was calculated from mathematical equations presented by Soudani et al. (2006) to the Ikonos images, and the calibration factor used to the bands was obtained from the Ikonos Soudani et al. (2006) and Goward et al. (2003).

The atmospheric correction, that aims the conversion of DN's (digital numbers) or brightness values from the original image into apparent reflectance was performed based on the empirical method of Chavez (1989), called DOS—Dark Object Subtraction, using the Application IDRISI Andes (Clark University). Among the parameters needed to implement the method in Idrisi, there is the value of Dn haze (digital number of the darkest pixel), which was achieved by the observation of the histograms frequencies to the original band which was being

corrected. It was stipulated that the value of Dn haze would be the one from the DL previous of the most abrupt frequency change observed in the histograms. However, there were observed occurrences of pixels with brightness values or frequencies already in DN = 1 to all bands of Ikonos images. Therefore, the value of Dn haze adopted was zero.

The resultant image from the georeferencing, from radiometric calibration and atmospheric correction was submitted then to a softening by the application of filters from simple moving average, in order to reduce the possible presence of noise and radiometric oscillations (Tsai and Philpot 1998). This process was carried out by IDRISI Andes application, with different sizes of windows and noticed that the results of correlation with higher values (bigger than 0.6) at a significance level of 5 %, took place in the filter window of 3×3 pixels. Performed the pre-processing operations from the image, the band ratio was applied to the values extracted from the points of concern, to the same intervals and combinations indicated by Kirk (1994), Barbosa (2005), Rundquist et al. (1996), Hoge et al. (1987) and Gitelson (1992), implying the four Ikonos bands:

$$(B3/B2) \tag{1}$$

$$(B2/B3) \tag{2}$$

$$[(B2 - B3)/(B2 + B3)] \tag{3}$$

$$[(B2 - B4)/(B2 + B4)] \tag{4}$$

$$(B4/B3) \tag{5}$$

$$(B3/B1) \tag{6}$$

$$(B4/B1) \tag{7}$$

$$(B3/B4) \tag{8}$$

$$(B1/B2) \tag{9}$$

$$(B1/B4) \tag{10}$$

First, the correlation analysis was performed between the spectral data and Secchi depth limnological variant, both collected in field. The correlations analysis was tried considering the four data groups in which the sample was divided, according to the weather conditions previously discussed, however, the group 1, 3 points were not included in the analysis because there was an insufficient number of elements, and also no significant correlation.

The groups 2, 3 and 4 (n = 10, n = 12 and n = 10 respectively) were included on the analysis with significant correlation.

The correlation coefficients obtained to the four groups from Secchi depth, applied the reason and difference bands in spectral intervals (with logarithmic function) is presented in Fig. 4.
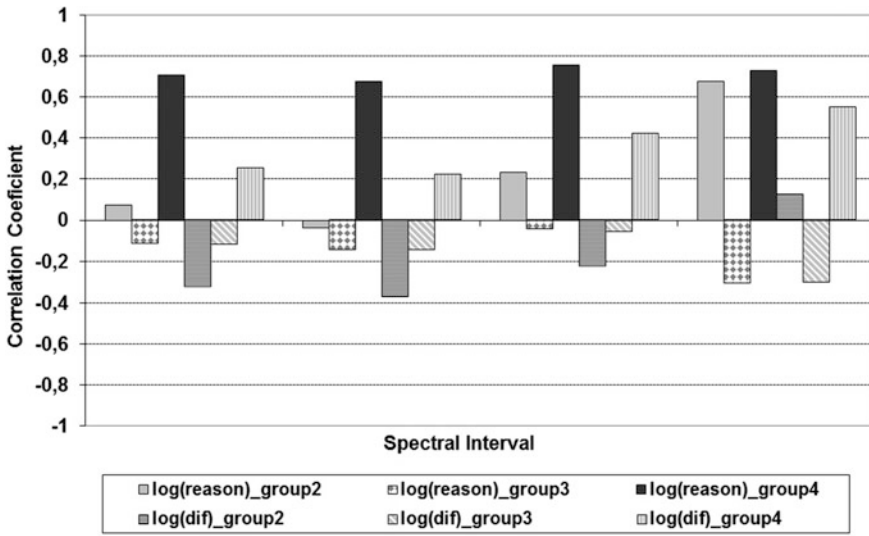
Fig. 4 Correlation coefficients between Secchi depth and logarithmic of ratio—log(reason) and differences—log(dif) for spectral intervals 571–698; 571–696; 568–648; 574–584 nm, to four significant groups

In Fig. 4, it can be noticed that Secchi depth is correlated by the significant level of 5 % between log(reason) and the band 574 and 584 nm (0.67; p = 0.032) to group 2. To other spectral intervals the correlations were lowest (<0.4) with no significant correlation at the 5 % level. Therefore, on group 2, it can inferred that band intervals of the reason from 574 to 584 nm were the best to detect the water transparence. To group 3, there is no significant correlation at the 5 % level, which were lower, what can conclude that clouds interfered negatively on the points. In Fig. 4, it can also be noticed that Secchi depth presented significant correlations at the significant level of 5 % to all band rate researched (Eqs. 11–14) in group 4. Thus, in the band difference, no significant correlation at the significant level of 5 %was noticed.

$$\left(\frac{571}{698}\right) \text{nm}_{(0.70; p = 0.022)} \tag{11}$$

$$\left(\frac{571}{696}\right) \text{nm}_{(0.67; p = 0.032)} \tag{12}$$

$$\left(\frac{568}{648}\right) \text{nm}_{(0.75; p = 0.012)} \tag{13}$$

$$\left(\frac{574}{584}\right) \text{nm}_{(0.73; p = 0.016)} \tag{14}$$

The correlation analyses were also performed between the image bands and Secchi depth limnological variant. It was an attempt to analyze the correlations considering the four data groups in which the sample was divided, according to the weather conditions previously discussed. However, not all of the points could be read on the multispectral image of February, due to the presence of clouds.

Consequently, group 1, with 3 points (P30, P32 and P36) and group 2 with 10 points (P22, P23, P25, P26, P27, P28, P29, P33, P34 and P35) were not included on the analysis because their points were under the clouds. Group 2 presented two spots whose reflectance values could be read (P34 and P35) and which were used as reference, due to the fact that it was impossible to apply the correlation tendency on them (n = 2 suggests no normality of data). Therefore, the analysis was leaded by making use of the group 3, with n = 11 points (P02, P05, P06, P07, P08, P09, P10, P12, P15, P16 and P19), and point-31 dropped for having been under the clouds, and the group 4 with 10 points (P01, P03, P14, P17, P18, P21, P104, P111, P117 and P129).

The correlation coefficients obtained from Ikonos multispectral imagebands and the Secchi depth variant are presented in groups 3 and 4 (Fig. 5a) and the correlation coefficients between Secchi depth and the band rate from Ikonos multispectral image in the groups 3 and 4 (Fig. 5b) tested on Eqs. 15–24.

$$R1 = \left(\frac{B3}{B2}\right) \tag{15}$$

$$R2 = \left(\frac{(B2 - B3)}{(B2 + B3)}\right) \tag{16}$$

$$R3 = \left(\frac{(B2 - B4)}{(B2 + B4)}\right) \tag{17}$$

$$R4 = \left(\frac{B4}{B3}\right) \tag{18}$$

$$R5 = \left(\frac{B3}{B1}\right) \tag{19}$$

$$R6 = \left(\frac{B4}{B1}\right) \tag{20}$$

$$R7 = \left(\frac{B3}{B4}\right) \tag{21}$$

$$R8 = \left(\frac{B1}{B2}\right) \tag{22}$$
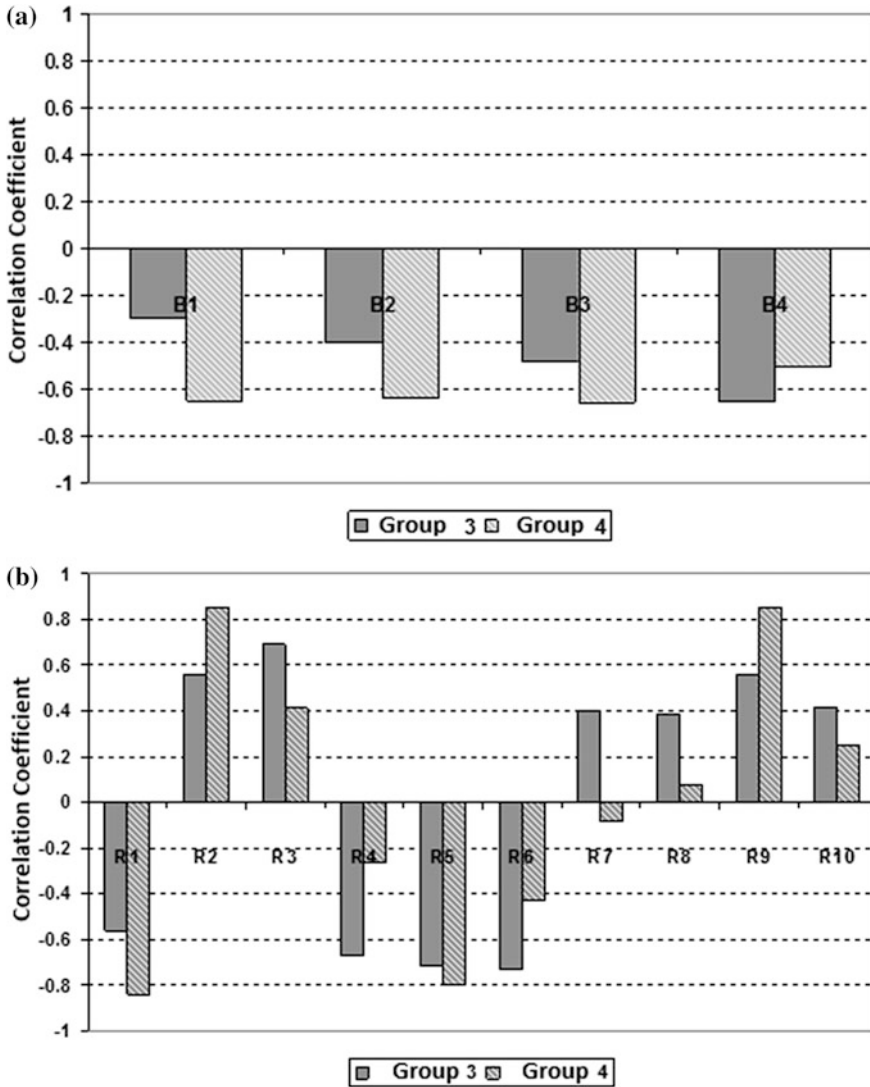
Fig. 5 **a** Correlation coefficients between Secchi depth and Ikonos multispectral bands to groups 3 and 4. **b** Correlation coefficients between Secchi depth and band rate from Ikonos multispectral imagery to groups 3 and 4 (Eqs. 15–24)

$$R9 = \left(\frac{B2}{B3}\right) \tag{23}$$

$$R10 = \left(\frac{B1}{B4}\right) \tag{24}$$

On the Fig. 5a it is noticed that to group 3 there was a significant correlation at the level 5 % only between the Secchi depth and band (B4) (-0.648; p $= 0.031$). To group 4 it is observed that there was a significant correlation at the level 5 % to bands (B1) (-0.649; p $= 0.042$); (B2) (-0.636; p $= 0.048$); and (B3) (-0.656; p $= 0.039$), which was no significant for B4. In Fig. 5b it is observed that the best correlation was to group 3 (Eq. 25) and having also higher correlation to Eq. 26 and lesser correlation to Eqs. 27 and 28.

$$R6 = \left(\frac{B4}{B1}\right) \text{with } r = -0.728; \ p = 0.011 \tag{25}$$

$$R5 = \left(\frac{B3}{B1}\right) \text{with } r = -0.716; \ p = 0.013 \tag{26}$$

$$R3 = \left(\frac{(B2 - B4)}{(B2 + B4)}\right) \text{with } r = -0.692; \ p = 0.018 \tag{27}$$

$$R4 = \left(\frac{B4}{B3}\right) \text{with } r = -0.67; \ p = 0.024 \tag{28}$$

To group 4 it is noticed higher correlations for ratio bands presented in Eqs. 29–32, hereafter:

$$R1 = \left(\frac{B3}{B2}\right)(-0.842; \ p = 0.002) \tag{29}$$

$$R2 = \left(\frac{(B2 - B3)}{(B2 + B3)}\right)(0.850; \ p = 0.002) \tag{30}$$

$$R5 = \left(\frac{B3}{B1}\right)(-0.795; \ p = 0.006) \tag{31}$$

$$R9 = \left(\frac{B2}{B3}\right)(0.854; \ p = 0.002) \tag{32}$$

4 Results

The analyses performed aiming the construction of an empiric inference model pointing to water transparence in the Itupararanga's/SP/Br Reservoir were made in groups 3 and 4.

Among the possible models, the one that seemed to be the most adequate statistically, by the defined criteria from Charnet et al. (1999) and Crusco et al. (2005), was that which describes the Secchi depth in function of the band rates

according to the Ikonos equation 20: (B4/B1) with R^2 adjusted explaining 47.8 % of the data variability, σ^2 within the average values from every models and Cp_Mallows adequated with Cp = 0.9 $\sim$ p = 1. The formal (Anderson–Darling) Normality Test from the dependent variant Secchi depth resulted positive to normality (A.D. = 0.316 e p-valor = 0.490), at the significance level of 5 %; beyond the p value resulted of the analysis from the variance of regression (0.011). Then, the model got defined by the following equation:

$$\text{Secchi Depth} = 2.74 - 1.14 \times \left(\frac{B4}{B1}\right) \tag{33}$$

The residue graphics from the model show their normality and a constant variance. The result from Anderson–Darling test confirms the result of normality (A.D. = 0.350 and p-value = 0.402), being then the indicated model to water-transparency inference according to the Secchi depth, for the group 3 data.

Although in group 4 there were significant correlation for Secchi depth and its spectral data and also Ikonos ratio bands, the regression models developed do not satisfied the statistical criteria by normality test of residuals and constant variance. Therefore it was not possible to produce inference models for group 4.

The validation from the group 3 regression model was accomplished by adapting the method of multiple sub-samples, known as "jackknife", which can be applied to small sizes samples, allowing the use of all observations in the parameter model estimation (Neophytou et al. 2000).

The method consists in separating an observation from the original sample, estimating the coefficients of the model based on the rest of the sample (n−1) and estimating the restricted observation using the new equation. The procedure is repeated to the entirety of the samples, as the observations could be estimated by models which parameters were estimated based on the others. The perceptual of right classifications is accumulated to all the sampling elements, indicating the global precision from the model.

The Table 2 presents the Interval of Prediction (IP) in the Secchi depth estimative for validation models with the application of the adapted method "jackknife".

On Table 2 it is verified that the real values of Secchi depth to all the sampling elements are included in the Intervals of Prediction (IP) from the models. In order to a more robust acceptance of the original model, it was applied the analysis of the coefficients $\beta 0$ $\beta 1$ from each validation model, noting if such coefficients belonged to the confidence interval (CI) from the original model, considering $\alpha = 5$ %. Being the confidence intervals $\beta 0$ and $\beta 1$ expressed by $-2.5426 < \beta 0 < 2.9373$ and $0.3299 < \beta 1 < 1.9500$, it was verified that all coefficients $\beta 0$ and $\beta 1$ from validation models are within the confidence interval (CI) from the original model.

Therefore, the original model is accepted, as the hypothesis of validation from the regression model developed from the Ikonos bands ratio (B4/B1) in the inference of Secchi depth is confirmed.

Table 2 Interval of prediction (IP) for Secchi depth from validation models with the adapted method "Jackknife"

Element sample excluded	Model validation: Secchi = $\beta 0 - \beta 1 \times$ (B4/B1)	ρ (B4/B1)	Secchi real (m)	Lower limit IP	Upper limit IP
1	Secchi = 2.7251 – 1.1227 × (B4/B1)	0.167	2.6	2.1823	2.8926
2	Secchi = 2.7268 – 1.1236 × (B4/B1)	0.154	2.6	2.1949	2.9122
3	Secchi = 2.7726 – 1.2694 × (B4/B1)	0.046	2.6	2.3256	3.1025
4	Secchi = 2.6816 – 0.9935 × (B4/B1)	0.120	2.8	2.2496	2.8758
5	Secchi = 2.7220 – 1.1959 × (B4/B1)	0.244	2.7	2.1490	2.7122
6	Secchi = 2.7756 – 1.2452 × (B4/B1)	0138	2.4	2.2805	2.9275
7	Secchi = 2.7343 – 1.1420 × (B4/B1)	0.212	2.5	2.1366	2.8484
8	Secchi = 2.7405 – 1.1338 × (B4/B1)	0.229	2.4	2.1306	2.8309
9	Secchi = 2.7573 – 1.1821 × (B4/B1)	0.174	2.4	2.2155	2.8876
10	Secchi = 2.7346 – 1.1397 × (B4/B1)	0.468	2.2	1.7109	2.6918
11	Secchi = 2.7148 – 0.9984 × (B4/B1)	0.399	22	1.9156	2.7168

In the model, it can be noticed that Ikonos multispectral ratio bands (B4/B1) modeled the Secchi depth concentrations in Itupararanga's Reservoir water, although this variable is less concentrated in its water body, as seen in Table 1. The empiric model described in Eq. 33 was applied at Ikonos image, using Idrisi software and is presented in Fig. 6.

Analyzing the mapping of Secchi depth, presented in Fig. 6, according to its variability in concentration by the mathematical model that uses bands B4 and B1 of Ikonos image from February 2007, it can be verified that these concentrations show homogeneous values in the entire Reservoir (between 1.83 and 2.65 nm)showing slightly higher values (more than 2.65 m)in central and west regions from the Reservoir. The field data (Table 1) presents Secchi depth values to points belonging to group 3 (between 2 and 2.80 m) in the west region from the Reservoir (Fig. 1). These tests show that the result of applying the empiric model is consistent with the Secchi depth values and characteristics from the water body observed on field, even with the presence of clouds in Ikonos images.
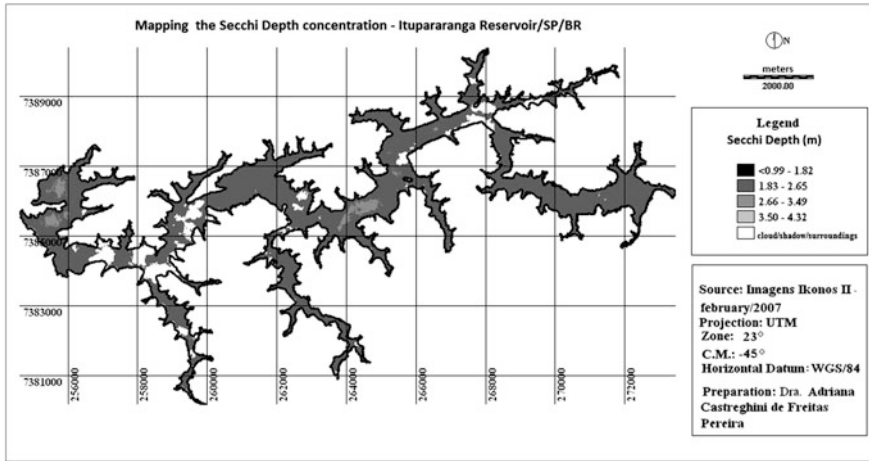
Fig. 6 Spatial distribution of inference model—mapping the Secchi depth concentration from B4/B1 Ikonos images

5 Conclusion

The main objective of the research was achieved, by developing the mapping of Secchi depth concentrations in Itupararanga's/SP/Br Reservoir water, from multispectral image from high spatial resolution Ikonos, spectral data and Secchi depth collected in situ; providing useful information to environmental experts. The empiric model established used the Ikonos image ratio band (B4/B1), corresponding to spectral bands 450–520 and 760–900 nm, respectively. The spatial distribution of inference model was consistent with the Secchi depth values and characteristics from the water body observed on field, even with the presence of clouds in Ikonos images.

It can be verified that the Secchi depth concentrations show homogeneous values in the entire Reservoir (between 1.83 and 2.65 nm)showing slightly higher values (more than 2.65 m)in central and west regions from the Reservoir. The field data presents Secchi depth values to points belonging to group 3 (between 2 and 2.80 m) in the west region from the Reservoir.

Economically, the research presented a good alternative to evaluate the water quality in reservoirs, replacing work field and lab analysis that, depending on the number of samples, make the research expensive and slow.

Generally, the total (economic) cost of the research was estimated in about 8.000 €, considered as low cost for water quality analysis in reservoirs, which makes the research economically viable.

The biggest obstacle faced by in the research was to adequate the obtainment of the Ikonos image and the collection of turbidity sampling in field, which theoretically should occur on the same day. Actually, due to the conditions of Ikonos

image programming, it occurred with the lag of 2 and 3 days from the field work and with considerable cloud presence, including over parts of important sites in the water from the Reservoir. Even then, the using of this image's bands was possible in the construction of the inference model of water turbidity, with good results.

6 Thanks

References

Barbosa CCF (2005) Sensoriamento Remoto da dinâmica da circulação da água do sistema planície de Curuai/Rio Amazonas. São José dos Campos. 281 p. Tese (Doutorado em Sensoriamento Remoto)—INPE

Charnet R, Freire CA De L, Charnet EMR, Bonvino H (1999) Análise de modelos de regressão linear com aplicações. Editora da Unicamp, Campinas, 356 pp

Chavez PS Jr (1989) Radiometric calibration of Landsat thematic mapper multispectral images. Photogramm Eng Remote Sens 55:1285–1294

Crusco NA, Anjos CS, Freitas CC, Renno CD, Epiphanio JCN (2005) Análise de regressão linear múltipla para simulação da banda do SWIR com outras bandas espectrais. In: XII Simpósio Brasileiro de Sensoriamento Remoto, 2005, Goiânia. Anais do XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, pp 891–898

Dekker AG (1993) Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing, 211 f. Tese (PhD theses), Free University, Amsterdam

Ekercin S (2007) Water quality retrievals from high resolution Ikonos multispectral imagery: a case study in Istanbul, Turkey. Water Air Soil Pollut 183:239–251

Esteves FA (1998) Fundamentos de limnologia, Rio de Janeiro: Interciência/Finep. 575 pp

Galvão LS, Formaggio AR, Tisot DA (2005) Discrimination of sugarcane varieties in southeastern Brazil with EO-1 Hyperion data. Remote Sens Environ 94(4):523–534

Giardino C, Pepe M, Brivio PA, Ghezzi P, Zilioli E (2001) Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. Sci Total Environ 268:19–29

Gitelson A (1992) The peak near 700nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. Int J Remote Sens 13(17):3367–3373

Goward SN, Davis PE, Fleming D, Miller L, Townshend JR (2003) Empirical comparison of Landsat 7 and Ikonos multispectral measurements for selected earth observation system (EOS) validation sites. Remote Sens Environ 88:80–99

Hoge EF, Wright CW, Swift RN (1987) Radiance ratio algorithm wavelengths for remote oceanic chlorophyll determination. Appl Opt 26(11):2082–2094

Kirk JTO (1994) Light and photosynthesis in aquatic ecosystems. Cambridge University Press, Cambridge, 509 pp

Louchard EM et al (2002) Derivative analysis of absorption features in hyperspectral remote sensing data of carbonate sediments. Opt Express 10(26):1573

Neophytou E, Charitou A, Charalambous C (2000) Predicting corporate failure: empirical evidence for the UK. Working Paper, University of Southampton

Nobrega IW (2002) Análise espectral de sistemas aquáticos da amazônia para a identificação de componentes opticamente ativos. São José dos Campos. 87 pp. Dissertação (Mestrado em Sensoriamento Remoto) Instituto Nacional de Pesquisas Espaciais (INPE)

Novo EMLM, Barbosa CCF, Freitas RM, Shimabukuro YE, Melack JM, Pereira Filho W (2006) Seasonal changes in chlorophyll distributions in Amazon floodplain lakes derived from MODIS images. Limnology 7(3):153–161

Pereira ACF (2008) Desenvolvimento de Método para Inferência de Características Físicas da Água associadas às variações espectrais. Caso de Estudo: Reservatório de Itupararanga/SP. Presidente Prudente, 206 pp. Tese (Doutorado em Ciências Cartográficas)—UNESP, Faculdade de Ciências e Tecnologia

Pereira Filho W (2000) Influência dos diferentes tipos de uso da terra em bacias hidrográficas sobre sistemas aquáticos da margem. Tese (Doutorado em Geografia)—USP

Pereira ACF, Galo MLBT, Velini E, Novo EMLM (2007). Amostragem em corpos d'água: Definição de elementos amostrais, posicionamento e coleta de dados "in situ". In: II Simpósio Brasileiro de Geomática e V Colóquio Brasileiro de Ciências Geodésicas. 2007, Presidente Prudente. Anais do II Simpósio Brasileiro de Geomática e V Colóquio Brasileiro de Ciências Geodésicas, pp 866–874

Pereira ACF, Galo MLBT, Velini E (2011). Inferência da Transparência da Água—Reservatório de Itupararanga/SP, a partir de imagens multiespectrais Ikonos e espectrorradiometria de campo. In: Revista Brasileira de Cartografia—RBC. On line

Ritchie JC, Cooper CM (1998) Comparison of measured suspended sediment concentration with suspended sediment concentrations estimated from Landsat MSS data. Int J Remote Sens 9(3):379–387

Rudorff CM (2006) Estudo da composição das águas da planície amazônica por meio de dados de reflectância do sensor Hyperion/eo-1 e de espectrorradiômetro de campo visando à compreensão da variação temporal dos seus constituintes opticamente ativos. 138f. Dissertação (Mestrado em Sensoriamento Remoto) Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos

Rundquist DC, Luoheng H, Schalles JF, Peake JS (1996) Remote measurement of algal chlorophyll in surface waters: the case for first derivative of reflectance near 690 nm. Photogramm Eng Remote Sens 62(2):195–200

Soudani K, François C, Le Maire G, Le Dantec V, Dufrêne E (2006) Comparative analysis of Ikonos, spot, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. Remote Sens Environ 102:161–175

Tsai F, Philpot W (1998) Derivative analyses of hyperspectral data. Remote Sens Environ 66:41–51

Landslide Susceptibility Mapping Along the National Road 32 of Vietnam Using GIS-Based J48 Decision Tree Classifier and Its Ensembles

Dieu Tien Bui, Tien Chung Ho, Inge Revhaug, Biswajeet Pradhan and Duy Ba Nguyen

Abstract The main objective of this study is to compare the results of decision tree classifier and its ensembles for landslide susceptibility assessment along the National Road 32 of Vietnam. First, a landslide inventory map with 262 landslide locations was constructed using data from various sources that accounts for landslides that occurred during the last 20 years. Second, ten landslide conditioning factors (slope, aspect, relief amplitude, topographic wetness index, toposhape, distance to roads, distance to rivers, distance to faults, lithology, and rainfall) were prepared. Third, using decision tree and two ensemble techniques i.e. Bagging and AdaBoost, landslide susceptibility maps were constructed. Finally, the resultant landslide susceptibility maps were validated and compared using a validation dataset not used during the model building. The results show that the decision tree with Bagging ensemble technique have the highest prediction capability (90.6 %), followed by the decision tree (87.8 %) and the decision tree with AdaBoost (86.2 %).

Keywords Decision tree · Ensemble technique · Landslide · GIS · Spatial analysis · Vietnam

D. Tien Bui (✉) · I. Revhaug
Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, Aas, Norway
e-mail: BuiTienDieu@gmail.comBui-Tien.Dieu@umb.no

Tien ChungHo
Department of Tectonic and Geomorphology, Vietnam Institute of Geosciences and Mineral Resources, Hanoi, Vietnam

B. Pradhan
Faculty of Engineering, Department of Civil Engineering, University Putra Malaysia, Kula Lumpur, Malaysia

D. Tien Bui · D. B. Nguyen
Faculty of Surveying and Mapping, Hanoi University of Mining and Geology, Hanoi, Vietnam

1 Introduction

Rainfall-triggered landslides are considered to be the most significant natural hazards in the north-western mountainous region of Vietnam (Tien Bui et al. 2012b). They have caused different types of damage affecting people, organizations, infrastructure, and the environment. The identification of areas susceptible to landslides is an essential task for assessing the landslide risk, and will contribute to public safety and decision-making in land management (Gorsevski et al. 2006). However, only a few landslide studies have been carried out and thus study of landslides is an urgent task in Vietnam (Tien Bui et al. 2012b).

Over the years, various methods and techniques for landslides prediction have been proposed and they vary from simple expert-based procedures to sophisticated mathematical models (Chung and Fabbri 2008). Review of these methods and techniques can be seen in Chacon et al. (2006).

Since the quality of landslide susceptibility models influence the method used to produce them (Yilmaz 2010), the investigation of new methods and techniques therefore, is higly necessary. In recent years, artificial intelligence techniques and data mining approaches are used in landslide studies and in general they outperform the conventional methods (Pradhan et al. 2010). In more recent years, ensemble-based approaches have received much attention in many fields including landslide studies. This is because the ensemble-based approaches have a capability to improve the prediction performance of models (Rokach 2010). In the ensemble methods, multiple classifiers are integrated and combined to produce the final model.

The main objective of this study is to apply the decision tree and its ensemble techniques for landslide susceptibility assessment along the National Road 32 of Vietnam. The difference between this study and the aforementioned literature is that two ensembles techniques: Bagging and AdaBoost were used. The computation process was carried out using MATLAB 7.11 and WEKA ver.3.6.6. Finally, a comparison of the results were made to choose the best one.

2 Study Area and Geospatial Database

The study area is along the corridor of the National Road 32 located in the north-western region of Vietnam. The total length of the road is 250 km (Fig. 1). The study area covers an area of about 3,164 km^2, between longitudes 103°33′23″E and 104°52′58″E, and between the latitude 22°20′18″N and 21°19′53″N.

Altitude ranges from 120 to 3,140 m a.s.l and decreases from the northwest to the southeast. The average altitude value is 1,078 m and standard deviation is 555.9 m. About 22.3 % of the total study area falls within slope group 0–15°; approximately 52.9 % of the study area has slope greater than 25°, and the remaining areas are in the slope category 15–25°.

More than 32 lithologic formations outcrop in this region and five of them (Suoi Bang, Muong Trai, Pu Tra, Ban Nguon, and Bac Son) cover together about 72 %
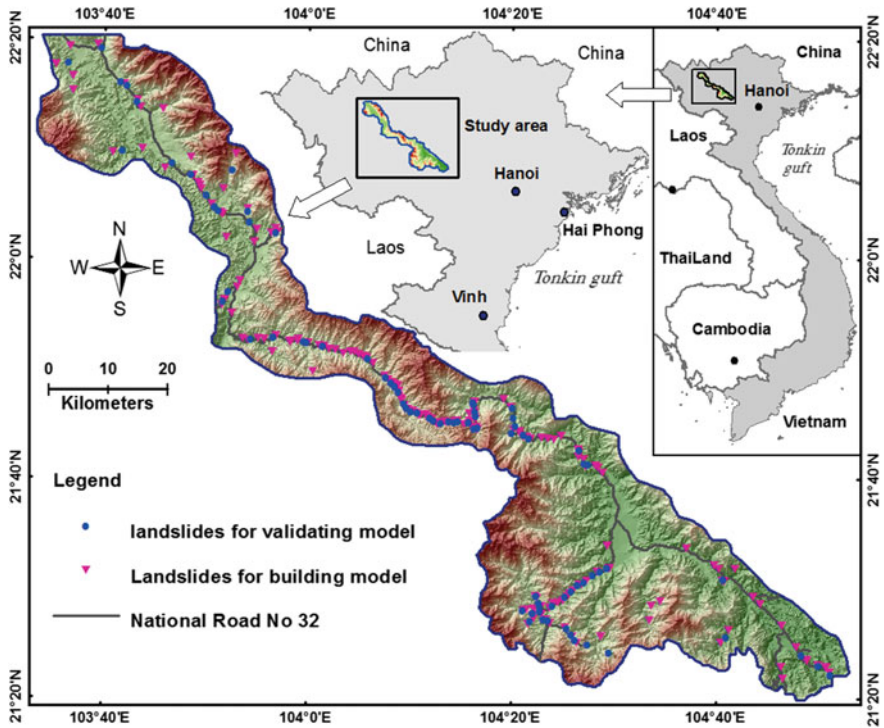
Fig. 1 Landslide inventory map of the study area

of the study area. The main lithologies are sandstone, conglomerate, clay shale, clayey limestone, siltstone, limestone, and clayey limestone.

In the study area, landslide locations were derived from the inventory map compiled earlier by Ho (2008). The landslide inventory (Fig. 1) was used to derive the quantitative relationships between the landslide occurrences and conditioning factors. A total of 262 landslides depicted by polygons were registered. These landslides occurred during the last 20 years. The smallest landslide size is about 476 m^2, the largest landslide size is 37,326 m^2.

In order to predict the location of future landslides, ten landslide conditioning factors were consideredin this study. Slope, aspect, relief amplitude, toposhape, and topographic wetness index (TWI) (Figs. 3, 4) were extracted from a digital elevation model (DEM) that was generated from national topographic maps at 1:50,000 scale. The resolution of the DEM is 20 m.

Distance to roads and distance to rivers maps (Figs. 5a, b) were constructed based on the river and road networks from the national topographic maps. Lithology map (Fig. 2) and distance to faults map (Fig. 4c) were constructed from the Geological and Mineral Resources Maps 1:200,000 scale. Finally, a rainfall map (Fig. 5c) was included in the analysis. The detailed classes for the ten landslide conditioning factors are shown in Table 1.
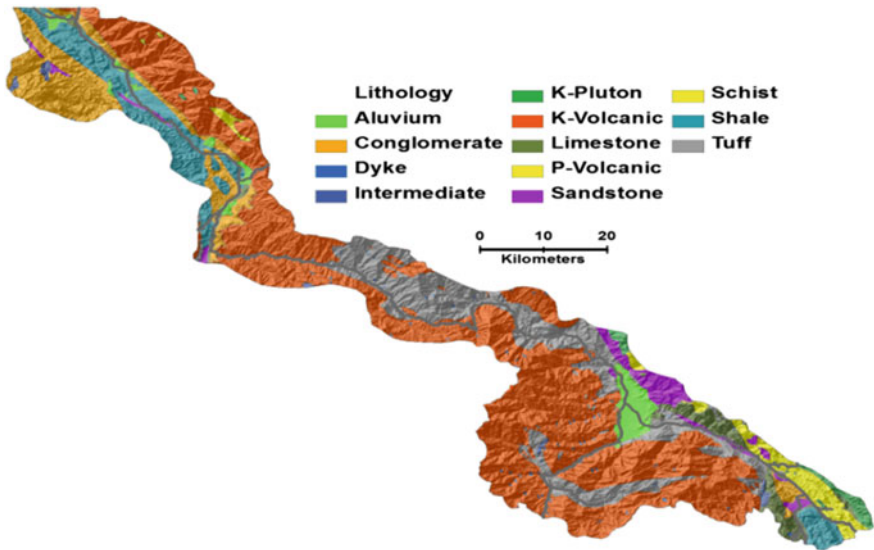
Fig. 2 Lithology map
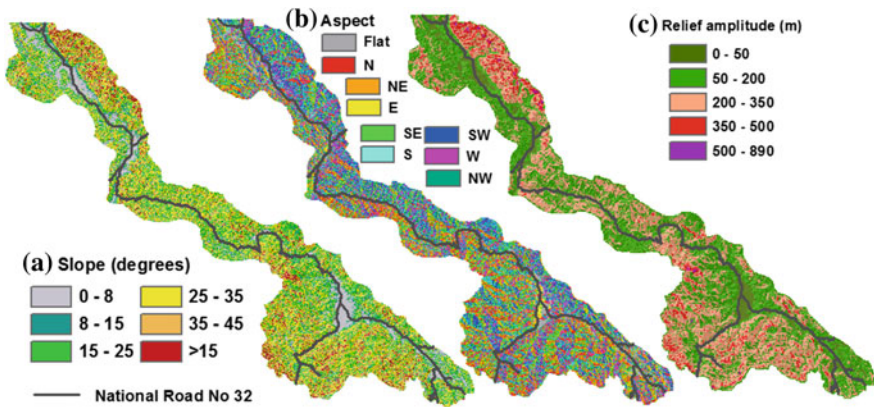
Fig. 3 **a** slope map; **b** aspect map; **c** relief amplitude map

3 Landslide Susceptibility Assessment Using the Decision Tree and Its Ensembles

3.1 Preparation of Training and Validation Data

The landslide inventory and ten conditioning factor maps were converted to a grid cell format with resolution of 20 m. Each category of the ten conditioning factor maps was assigned an attribute value and then was normalized to the range 0.1–0.9
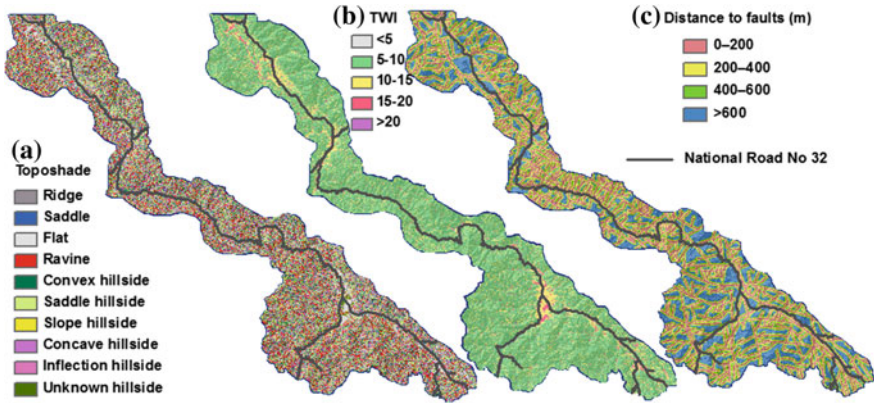
Fig. 4 a Toposhade map; **b** topographic wetness index (TWI) map; **c** distance to faults map
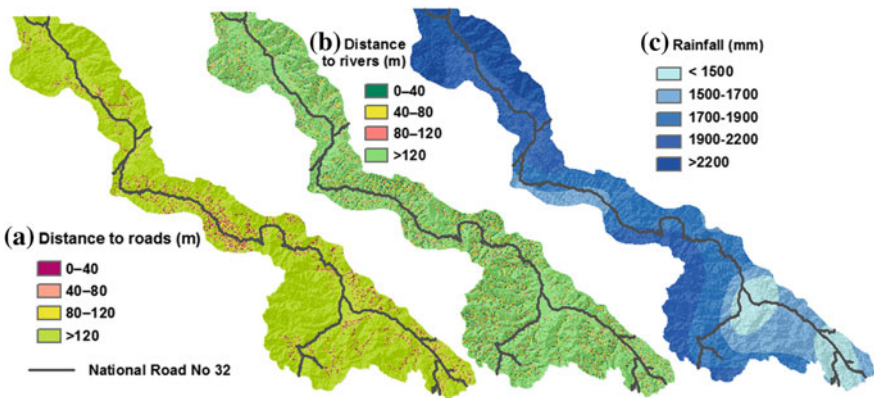
Fig. 5 a Distance to roads map; **b** distance to rivers map; **c** rainfall map

(Table 1) using the Max–Min formula (Tien Bui et al. 2012a). The attribute value was obtained based on frequency ratio. In the landslide inventory map, a "1" was assigned to landslide pixels whereas a "0" was assigned for pixels outside a landslide i.e. non-landslide pixels.

To evaluate the prediction capability of a landslide model, a landslide inventory should be split into two subsets; one is used for training and the other is used for validation (Chung and Fabbri 2003). Since the dates for the past landslide are unknown, the temporal division of the landslide inventory map is impossible. In this study, the landslide inventory map was randomly split in a 70/30 ratio for training and validation of the model, respectively (Fig. 1). In the next step, a total of 2,781 non-landslide pixels were randomly sampled from the landslide-free area. Finally, values for the ten conditioning factors were then extracted to build a training dataset.

Table 1 Landslide conditioning factors and their classes

Data layers	Class	Class Pixel	Landslide Pixel	Frequency Ratio	Attribute	Normalized Classes
Slope	0–8	1050626	0	0.000	1	0.10
(degree)	8–15	707774	197	0.578	2	0.45
	15–25	1949706	1226	1.305	3	0.90
	25–35	2352056	1471	1.298	4	0.89
	35–45	1379361	748	1.126	5	0.79
	>45	431672	150	0.721	6	0.54
Aspect	Flat	370810	0	0.000	1	0.10
	North	880893	282	0.665	2	0.39
	Northeast	954851	217	0.472	3	0.31
	East	887194	468	1.095	4	0.58
	Southeast	943832	595	1.309	5	0.68
	South	1016869	891	1.819	6	0.90
	Southwest	1061249	622	1.217	7	0.64
	West	893222	252	0.586	8	0.36
	Northwest	862275	465	1.119	9	0.59
Relief	0–50	494988	85	0.356	1	0.30
Amplitude	50–200	3797449	2664	1.456	2	0.90
	200–350	3106823	994	0.664	3	0.46
	350–500	449532	49	0.226	4	0.22
	>500	27992	0	0.000	5	0.10
TWI	<5	744	0	0.000	1	0.10
	5–10	6246498	3538	1.176	2	0.90
	10–15	1328468	249	0.389	3	0.36
	15–20	233435	5	0.044	4	0.13
	>20	20661	0	0.000	5	0.11
Toposhape	Ridge	1437448	744	1.074	1	0.81
	Saddle	113672	0	0.000	2	0.10
	Flat	374668	0	0.000	3	0.11
	Ravine	1399148	546	0.810	4	0.63
	Convex hillside	1030755	517	1.041	5	0.78
	Saddle hillside	2408283	1403	1.209	6	0.89
	Slope hillside	16366	0	0.000	7	0.12
	Concave hillside	945577	555	1.218	8	0.90
	Inflection hillside	60540	27	0.926	9	0.71
	Unknown hillside	90484	0	0.000	10	0.13
Lithology	Aluvium	239956	79	0.683	1	0.10
	Conglomerate	789689	188	0.494	2	0.11
	Dyke	27674	23	1.725	3	0.21
	Intermediate	42216	0	0.000	4	0.24
	K-Pluton	87073	0	0.000	5	0.30
	K-Volcanic	4030918	1725	0.888	6	0.38
	Limestone	240162	39	0.337	7	0.43
	P-Volcanic	5770	0	0.000	8	0.46
	Sandstone	237588	192	1.677	9	0.48

(continued)

Table 1 (continued)

Data layers	Class	Class Pixel	Landslide Pixel	Frequency Ratio	Attribute	Normalized Classes
	Schist	274344	125	0.946	10	0.78
	Shale	679581	269	0.822	11	0.80
	Tuff	1204892	1152	1.985	12	0.90
Distance	0–200	2419662	1636	1.403	1	0.90
to Faults	200–400	2027888	989	1.012	2	0.53
(m)	400–600	1442500	635	0.914	3	0.44
	>600	1986891	532	0.556	4	0.10
Distance to	0–40	273124	1139	8.656	1	0.90
Roads	40–80	292995	958	6.787	2	0.72
(m)	80–120	288433	582	4.188	3	0.47
	>120	7022389	1113	0.329	4	0.10
Distance to	0–40	541068	467	1.792	1	0.77
Rivers (m)	40–80	581557	518	1.849	2	0.80
	80–120	576604	555	1.998	3	0.90
	>120	6177712	2252	0.757	4	0.10
Rainfall	<1500	892649	215	0.500	1	0.18
(mm)	1500–1700	1397443	770	1.144	2	0.64
	1700–1900	2272315	1644	1.502	3	0.90
	1900–2200	2060447	925	0.932	4	0.49
	>2200	1254071	238	0.394	5	0.10

3.2 Decision Tree Classifier

A decision tree classifier represented by a tree-like structure is a hierarchical model composed of internal nodes, leaf nodes, and branches. The main advantage of decision trees is that they are easy to construct and the resulting trees are readily interpretable. The main disadvantage of decision trees is that multiple output are not allowed and decision trees are susceptible to noisy data (Zhao and Zhang 2008).

Various construction algorithms for decision trees have been proposed in the literature. For this study, the J48 algorithm that is a Java reimplementation of the C4.5 algorithm (Witten et al. 2011) was used. The detailed explanation of the C4.5 algorithm can be seen in Quinlan (1993).

Table 2 Parameters for the decision tree

Parameters	Selected
Type of pruning	Based on sub-tree raising
Confidence factor for tree pruning	0.2
Binary splits or multiple splits	Multiple splits
Minimum number of instances per leaf	8
Using Laplace smoothing	True

The first step in constructing the landslide model using decision trees is to determine parameters that influence the size of the result tree. Therefore, a test has been carried out to find the most suitable parameters for the study area. The most preferable parameters are selected based on the classification accuracy. The results are shown in Table 2.

Using the training data set and the determined parameters, the model was trained using the stratified tenfold cross-validation method. For each running, one fold was used for testing whereas the remaining ninefolds were used for training the model. Finally, the decision tree model for landslide susceptibility was constructed. The size of the tree is 189 including the root node, 93 internal nodes, and 95 leafs. The detail accuracy by class and performance of the decision tree model is shown in Tables 3, and 4.

3.3 Bagging and AdaBoost Ensemble Approaches

An ensemble decision-tree based classifier is defined as a classifier that combines multiple trained decision-tree classifiers to produce a single final classification.

Bagging is known as bootstrap aggregation, is one of the earliest ensemble algorithms proposed by Breiman (1996). Bagging uses bootstrap sampling, which is a random sampling with replacement, to generate multiple subsets from the training dataset. Then each of the subset is used to construct a decision-tree based

Table 3 Performance of the decision tree, the decision tree with Bagging, and the decision tree with AdaBoost

Model	True positive rate (%)	False positive rate (%)	F-measure (%)	Class
Decision tree	0.919	0.153	0.887	Landslide
	0.847	0.081	0.878	No-landslide
Decision tree with Bagging	0.925	0.130	0.900	Landslide
	0.870	0.075	0.895	No-landslide
Decision tree with AdaBoost	0.941	0.114	0.916	Landslide
	0.886	0.059	0.911	No-landslide

Table 4 Accuracy assessment by classes of the decision tree, the decision tree with Bagging, and the decision tree with AdaBoost

Parameters	Decision tree	Decision tree with Bagging	Decision tree with AdaBoost
Classification accuracy (%)	88.28	89.75	91.33
Cohen's Kappa index	0.765	0.795	0.827
Root mean squared error (RMSE)	0.307	0.286	0.273

model. A final decision-tree model is determined by aggregating all decision-tree based models.

AdaBoost known as adaptive boosting is a method that combines multiple base classifiers (Freund and Schapire 1997). The algorithm starts by constructing an initial decision-tree based model using a subset of the training dataset. In this step, the instances have equal weights. This model then predicts all instances in the training dataset. The misclassified instances will be assigned higher weights. The weights of the correctly classified instances are kept unaltered. In the next step, the weights of all instances in the whole training dataset are normalized and a new subset is then randomly sampled to build a next decision-tree based model. This process continues until it reaches a terminated condition. The final decision-tree model is obtained based on a weighted sum of all the decision-tree based models.

Using the training data set, the decision trees with Bagging and AdaBoost were constructed. A total of 30 iterations were used to create ensembles for the two ensemble approaches. The parameter setting for the decision tree remains the same as in Sect. 3.1. Model training was based on stratified 10-fold cross-validation. The results is shown in Tables 3, and 4.

3.4 Reclassification of Landslide Susceptibility Maps

The susceptibility indexes were reclassified into five different susceptibility classes based on the percentage of area (Pradhan and Lee 2010): very high (10 %), high (10 %), moderate (20 %), low (20 %), and very low (40 %) (Figs. 6, 7, and 8).

4 Validation and Comparison of the Landslide Susceptibility models

4.1 Model Evaluation

In order to assess the performance of the susceptibility models, several statistical evaluation criteria were used (Tien Bui et al. 2012a): (i) model accuracy; (ii) model sensitivity (true positive rate); (iii) model specificity (true negative rate); (iv) Cohen's kappa; (v) and the area under the success-rate curve (AUC).

The results of the performance evaluation of the three landslide model are shows in Tables 3, 4, and 5. It could be seen that all the models have a high classification accuracy. However, the decision tree with Bagging and Adaboost outperform the single decision tree model.

The Cohen's kappa indexes that measures the reliability of the susceptibility models for decision tree and decision tree with Bagging are 0.765 and 0.795, respectively. They indicate that a substantial agreement between the susceptibility
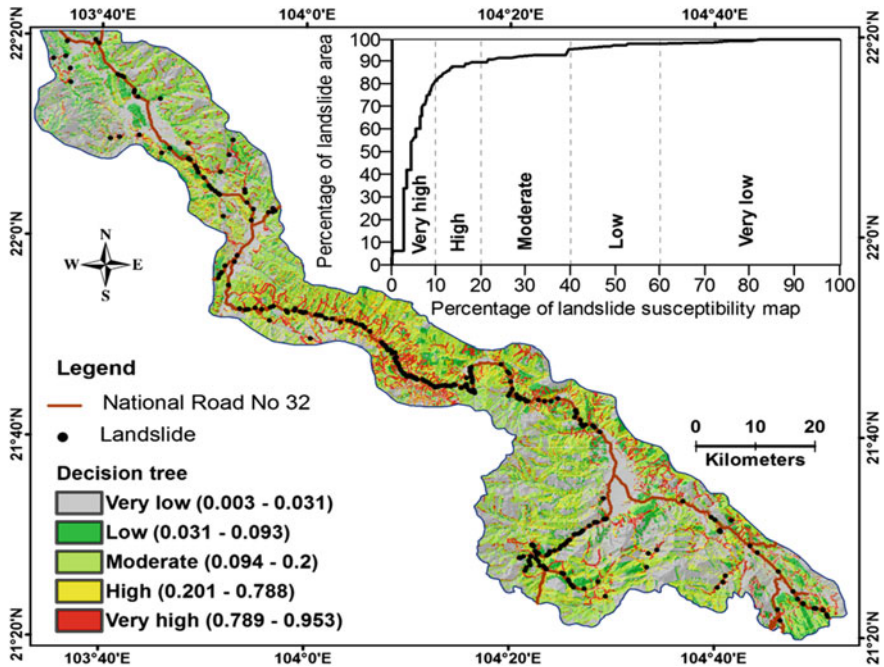
Fig. 6 Landslide susceptibility map using the decision tree

model and reality. For decision tree with Bagging, Cohen's kappa index is 0.827 indicates a good agreement between the susceptibility model and reality.

The detailed accuracy assessment by classes is shown in Table 4. It can be seen that the True Positive rates and the F-measures are higher for the landslide class than for the no-landslide class for all the three susceptibility models.

The success-rate curves for three models were derived by comparing the 2,781 landslide grid cells in the training dataset with the three susceptibility maps. The areas under the success-rate curves (AUC) were then estimated for all cases (Table 5). The result shows that all three models have a good fit with the training dataset. The highest degree of fit is for the decision trees with AdaBoost (0.955), followed by the decision tree with Bagging (0.933) and the single decision tree (0.916).

4.2 Prediction Rate

By comparing the landslide grid cells in the validation dataset with the three landslide susceptibility maps, three prediction-rate curves for the landslide models were obtained (Fig. 6).
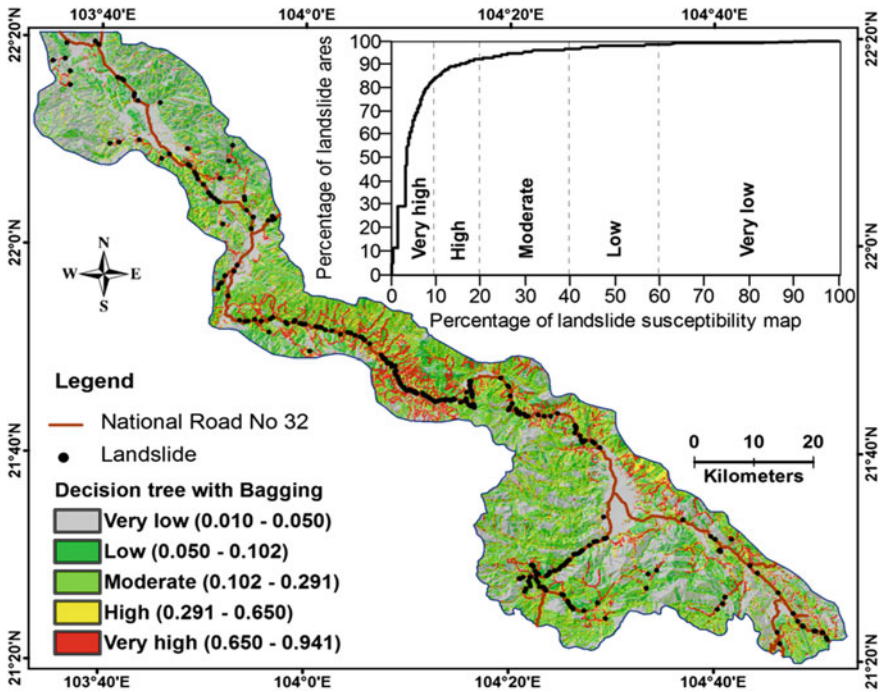
Fig. 7 Landslide susceptibility map using the decision tree with Bagging

The closer the curve is to the upper-left corner, the better is the model. For quantitative comparison, the areas under theprediction-rate curves (AUC) were further calculated. The closer the AUC value is to 1, the better is the model. The result (Fig. 9) show that the decision tree with Bagging has highest prediction capability of future landslides (AUC = 0.906), followed by the decision tree (AUC = 0.878) and the decision tree with AdaBoost (AUC = 0.862).

4.3 Relative Importance Assessment of the Conditioning Factors

The relative importance of the landslide conditioning factor in the susceptibility models was estimated by excluding each factor and then calculated the classification accuracy of the models (Table 6). It could be seen that distance to roads have the highest contribution to the three models. This is because the landslide locations were mainly in the corridor of the National Road No 32. Distance to faults, rainfall, lithology, slope, and relief amplitude have the highest contribution to all three models. In contrast, toposhade and TWI contribute less. For the case of
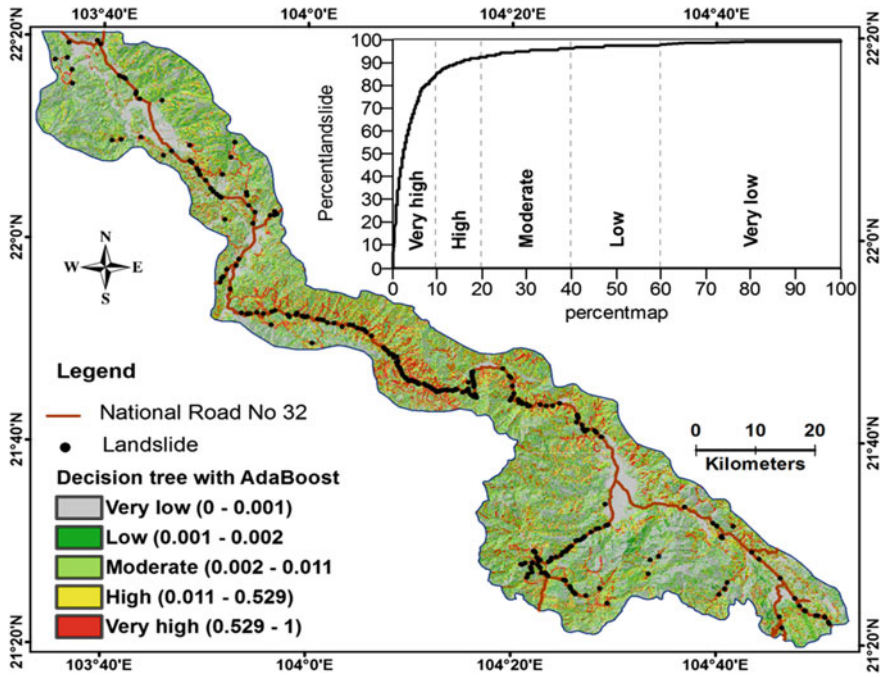
Fig. 8 Landslide susceptibility map using the decision tree with AdaBoost

Table 5 Success rate of the landslide susceptibility models

Susceptibility models	Area under the success-rate curve (AUC)
Decision tree	0.916
Decision tree with Bagging	0.933
Decision tree with AdaBoost	0.955

distance to rives, this factor has a contribution to the two ensemble models. However, it might have caused slightly noise by reducing the classification accuracy by 0.66 % in the decision tree model.

5 Discussions and Conclusion

In this study, the two ensemble-based (Bagging and AdaBoost) decision tree models were developed and applied for landslide susceptibility assessment along National Road No 32 (Vietnam). For comparison, a single decision tree model was also used. Using ten conditioning factors and the landslide locations in the training dataset, three susceptibility models were trained using ten-fold cross-validation
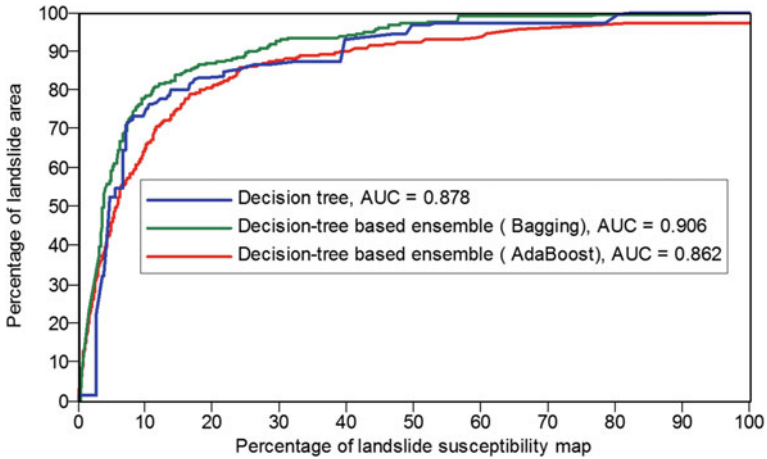
Fig. 9 Prediction-rate curves and area under the curves (AUC) for the decision tree, the decision tree with Bagging, and the decision tree with AdaBoost

Table 6 Relative importance of the landslide conditioning factors

Conditioning factors	Classification accuracy (%)		
	Decision tree	Decision tree with Bagging	Decision tree with AdaBoost
Minus slope	87.95	88.87	91.06
Minus aspect	86.98	87.39	88.20
Minus relief amplitude	87.55	88.9	90.07
Minus TWI	88.42	89.85	91.38
Minus toposhade	88.38	89.73	92.63
Minus lithology	87.36	88.17	88.71
Minus distance to faults	86.33	87.36	89.30
Minus distance to roads	81.01	83.08	85.33
Minus distance to rivers	88.94	89.21	90.85
Minus rainfall	87.07	87.95	88.44
All	88.28	89.75	91.33

method. The final models were then applied to construct three landslide susceptibility maps. These maps only represent spatial predictions of future landslides. They do not provide information about "when" and "how frequently" a landslide will occur.

The performance evaluation results show that the classification accuracy of the decision-tree models with Bagging and AdaBoost increased of about 1.47 % and 3.05 % respectively, compared to the single decision-tree model. The evaluation of the degrees of fit of the models with training dataset show that the decision-trees with Bagging and AdaBoost have slightly better compared to the singe decision tree model.

Using Cohen's Kappa index, the reliabilities of the landslide models were assessed. The index results from 0.765 to 0.827 show a substantial agreement between the susceptibility model and reality for all the models. The results are satisfying compared with other works such as Saito et al. (2009) and Tien Bui et al. (2012a).

The prediction capability of the susceptibility models were estimated using landslide location data that was not used in the training phase. The results show that the decision tree model with Bagging has the highest prediction capability. In the case of the decision tree model with AdaBoost, although this model has the highest degree of fit to the training data, the prediction capability has the lowest value.

The relative importance of the ten conditioning factors for the three susceptibility models show that distance to roads, distance to faults, rainfall, lithology, slope, and relief amplitude have a high contribution to all the models, the highest for the first one. This result is different compared to studies carried out by others such as Pradhan and Lee (2010) and Van Den Eeckhaut et al. (2006) where slope is indicated as the most important factor. The difference is due to the fact that this study only focuses on landslides along the corridor of the National Road 32 of Vietnam.

As a final conclusion, the finding of this results suggested that the decision tree model with Bagging is the most preferable in this study. The results may be useful for policy planning and decision making in areas prone to landslides.

References

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Chacon J, Irigaray C, Fernandez T, El Hamdouni R (2006) Engineering geology maps: landslides and geographical information systems. Bull Eng Geol Environ 65:341–411

Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. Nat Hazards 30:451–472

Chung C-J, Fabbri AG (2008) Predicting landslides for risk analysis—spatial models tested by a cross-validation technique. Geomorphology 94:438–452

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139

Gorsevski PV, Gessler PE, Boll J, Elliot WJ, Foltz RB (2006) Spatially and temporally distributed modeling of landslide susceptibility. Geomorphology 80:178–198

Ho TC (2008) Application of structural geology methods, remote sensing, and GIS for the assessment and prediction of landslide and flood along the National Road 32 in the Yen Bai and Lai Chau provinces of Vietnam. Vietnam Institute of Geosciences and Mineral Resources, Hanoi, p 118

Pradhan B, Lee S (2010) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. Environ Earth Sci 60:1037–1054

Pradhan B, Sezer EA, Gokceoglu C, Buchroithner MF (2010) Landslide susceptibility mapping by neuro-fuzzy approach in a landslide-prone area (Cameron Highlands, Malaysia). IEEE Trans Geosci Remote Sens 48:4164–4177

Quinlan JR (1993) C4.5: programs for machine learning Morgan Kaufmann San Mateo

Rokach L (2010) Ensemble-based classifiers. Artif Intell Rev 33:1–39

Saito H, Nakayama D, Matsuyama H (2009) Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: the Akaishi Mountains, Japan. Geomorphology 109:108–121

Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012a) Landslide susceptibility assessment in Vietnam using support vector machines, Decision tree and Naïve Bayes models. Mathl Probl Eng. doi:10.1155/2012/9746382012:26

Tien Bui D, Pradhan B, Lofman O, Revhaug I, Dick OB (2012b) Landslide susceptibility assessment in the Hoa Binh province of Vietnam: a comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. Geomorphology 171–172:12–29

Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006) Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76:392–410

Witten IH, Frank E, MA H (2011) Data mining: practical machine learning tools and techniques, 3rd edn. Morgan Kaufmann, Burlington

Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. Environ Earth Sci 61:821–836

Zhao Y, Zhang Y (2008) Comparison of decision tree methods for finding active objects. Adv Space Res 41:1955–1959

GIS-Based Landslide Susceptibility Mapping Using Remote Sensing Data and Machine Learning Methods

Fu Ren and Xueling Wu

Abstract In the Three Gorges of China, there are often landslide disasters, and the potential risk of landslides is tremendous. Thus, an efficient and accurate method of generating landslide susceptibility maps is very crucial to mitigate the loss of lives and properties caused by these landslides. This study presents a multidisciplinary approach to map landslide susceptibility on the Xietan Town of the Three Gorges, using slope units, intelligent models, geographic information system (GIS), and remote sensing data. Thirteen environmental factors, which have been extracted from 1:10,000-scale topographic maps, 1:50,000-scale geological maps, and HJ-1A satellite images with a spatial resolution of 30 m, were selected as predictor variables, including slope, aspect, curvature, slope unit altitude, engineering rock group, slope structure, distance from faults, land use, normalized difference vegetation index, reservoir water level, distance from drainage, catchment area, and catchment height. A two-class support vector machine (SVM) was trained and used to assess landslide susceptibility. Area under the curve was used to validate performance of the models. The results show that the two-class SVM outperforms the back propagation neural network in terms of both accuracy and generalization capacity, the area ratio being 0.8365 and approximately 90 % of landslides were classified as high and very high landslide-prone areas.

Keywords GIS · Landslide · Susceptibility · Support vector machine (SVM) · Three Gorges

F. Ren (✉)
School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China
e-mail: renfu@whu.edu.cn

X. Wu
Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China

1 Introduction

Landslide is defined as the movement of a mass of rock, debris, or earth (soil) down a slope (Fell et al. 2008). As one of the most common natural disasters, landslide endangers the lives and properties of nearby inhabitants. Landslide activity has an increasing trend worldwide, especially in the developing countries. This trend is expected to continue in the next decades because of the increased urbanization and development, continued deforestation, and increased regional precipitation in landslide-prone areas.

Among the measures that have been taken to avoid the development of threatened areas, the generation of landslide susceptibility maps is the most efficient way to reduce future damage and loss of lives. Landslide susceptibility map can provide valuable information to planners, developers, and engineers who implement land use strategies not only in the design stages but also in the hazard mitigation stages (Yilmaz et al. 2012). The main purpose of landslide susceptibility mapping is to identify areas of future mass movements, referring to the previous knowledge about the spatial distribution of past occurrences, assuming that landslides will be prone to occur in the future because of the same conditions that produced them in the past.

Landslide is an indeterminacy phenomenon. However, traditional spatial analytical techniques lack data mining of nonlinear landslide systems and cannot easily discover new and unexpected patterns, trends, and relationships which can be hidden deep within very large diverse geographic datasets. In recent years, with the rapid development of remote sensing (RS) and geographic information system (GIS), data sources of landslides have moved from a data-poor and computation-poor to a data-rich and computation-rich environment. Therefore, the current trend is to determine how to take full advantage of the large amount of available multi-source data to achieve a quantitative landslide susceptibility assessment with high accuracy and efficiency. At this point, extensive attention has been paid internationally to the study of introducing intelligent methods to landslide susceptibility mapping. Artificial neural network (ANN) was widely applied in landslide susceptibility assessment (Chauhan et al. 2010; Choi et al. 2010; Pradhan and Lee 2010; Pradhan et al. 2010; Zarea et al. 2012), but the ANN-based methods cannot provide objective and steady output because their outputs are operator-dependent. Decision tree is also an efficient tool for landslide susceptibility prediction (Yeon et al. 2010; Bui et al. 2012; Pradhan 2012), but it is limited in dealing with many layers of spatial data. Support vector machine (SVM), originally developed by Vapnik (1995), has become an increasingly popular method to produce landslide susceptibility maps (Yao et al. 2008; Bui et al. 2012; Ballabio and Sterlacchini 2012; Pradhan 2012). As an inherently robust and collinearity resistant technique, SVM can help to overcome some assumptions and limitations implied in the above models.

Moreover, landslide susceptibility mapping always poses the need to define a basic spatial mapping unit because it affects the reliability of the obtained susceptibility maps (Erener and Duzgun 2012; Rotigliano et al. 2012). The pixel-

oriented methods commonly used possess a great number of samples and their prediction results appear discontinuous to some degree, but slope unit-oriented methods need quite fewer samples and can produce continuous prediction results. Hence, slope units are selected as the basic mapping unit in this study.

In the Three Gorges of China, there are frequent landslides, and the potential risk of landslides is tremendous. Landslide susceptibility mapping is of great significance to guarantee the normal operation of the Three Gorges Dam and to safeguard the properties and lives of local residents. Hence, slope units were extracted from a digital elevation model (DEM), and thirteen thematic maps were prepared for landslide susceptibility mapping using GIS and image processing. Then, a two-class SVM was applied to landslide susceptibility mapping. The experimental result was visualized in GIS environment and validated by comparing with the existing landslides according to area under the curve (AUC).

2 Description of the Study Area

The study area is located in the west of Xiling Gorge, approximately between 30°59–31°02′N and 110°32′–110°39′E with an area of about 50 km^2, including Xietan Town (Zigui County) of Three Georges. Geomorphology of the area is characterized by a rugged topography with the hill ranges varying from 80 to 1220 m (Fig. 1). Yangtze River broadly crosses the study area in the NWW–SEE direction. Climate in the area is typical subtropical and monsoonal, having hot and humid summers but cold and dry winters, with 1006.8 mm annual average rainfall. There are twenty-eight large old landslides after construction of the reservoir, with an area of about 3.163 km^2, and the maximum and minimum areas are 0.854 and 0.008 km^2, respectively.

According to the lithological characteristics and engineering properties, the strata covering the Three Gorges were divided into four groups: I, II, III, and IV. II is further classified into three sub-regions. The study area mainly consists of II3, III, and IV (Fig. 2). The shales, marls, and peat intercalations along the river are liable to landslides.

3 Data Sources

One scene of HJ-1A satellite images (Path7/Row76) with a spatial resolution of 30 m, equipped with a CCD camera, acquired on 13th April 2011 by China Center for Resources Satellite Data and Application, was used in this study. The raw images were processed by ENVI 4.7 software.

A DEM with a spatial resolution of 28.5 m was generated from the 1:10,000-scale digital topographic maps, surveyed by Three Gorges Headquarters.

Fig. 1 Location of the study area
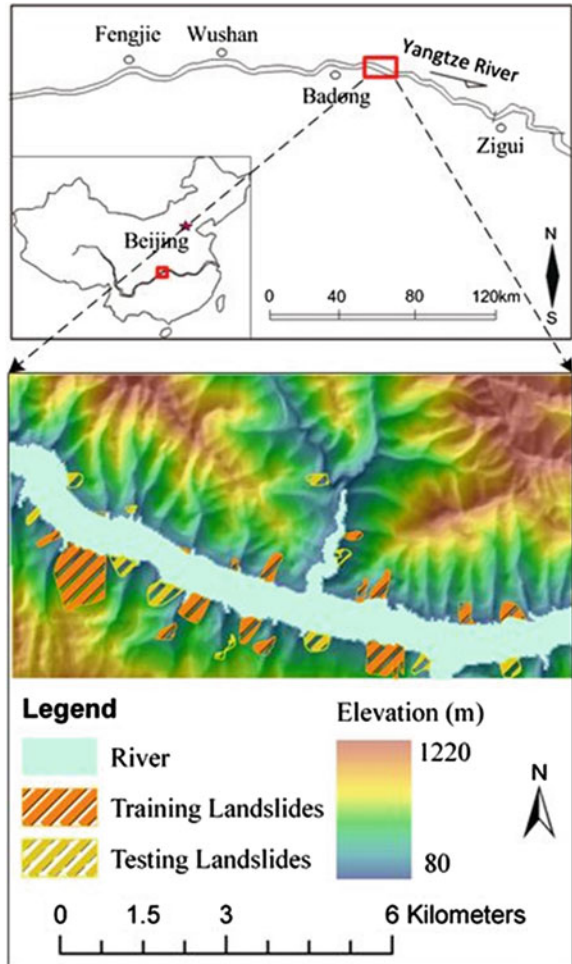
1:50,000-scale digital geological maps surveyed by Three Gorges Headquarters were used to obtain geological parameters. In addition, field investigation data were used to validate the accuracy of landslide susceptibility maps.

4 Methods

SVM is an increasingly popular learning procedure based on statistical learning theory. There are two main ideas underlying the SVM. One is an optimum linear hyperplane which is applied to separate the data patterns. The other is the use of kernel functions to convert original non-linear data patterns into a format that is
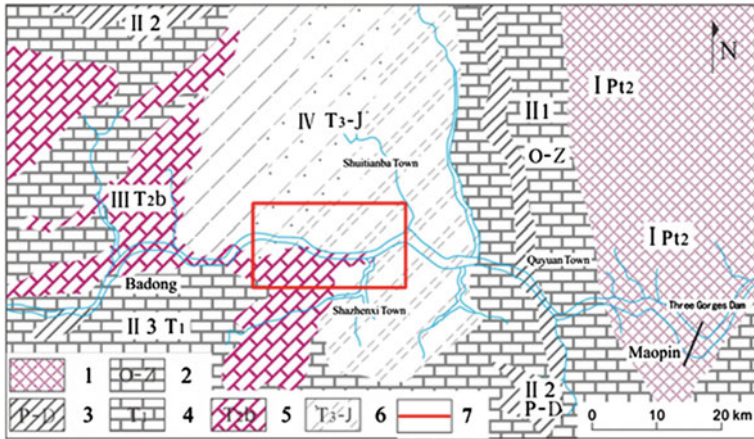
Fig. 2 Stratum zoning map of the Three Gorges (*1*-I Pre-Sinian crystalline basement; *2*-II1 Sinian-Ordovician limestone and dolomite; *3*-II2 Silurian-Triassic sandstone, siltstone and shale; *4*-II3 Lower Triassic limestone and dolomite; *5*-III T2b sandstones, shales and marls; *6*-IV interbedded sandstones and siltstones, marls and peat intercalations; *7*-Study area) (after Yangtze Water Resources Commission, 1992, modified)

linearly separable in a high dimensional feature space (Yao et al. 2008). In this study, a two-SVM was applied to generate landslide susceptibility map. The pertinent points of two-class SVM are presented herein. A set of linear separable training vectors $x_i (i = 1, 2, \ldots, N)$ is composed of two-class ($y_i = \pm 1$). The goal of SVM modeling is to find an n-dimensional hyperplane discriminating the maximum gap between landslide-prone and landslide-not-prone mapping units. When using the method of Lagrange multipliers λ_i, the cost function can be defined as

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{N} \lambda_i (y_i ((w \cdot x_i) + b) - 1) \tag{1}$$

and

$$\frac{1}{2} \|w\|^2 \tag{2}$$

subject to

$$y_i ((w \cdot x_i) + b) \geq 1 \tag{3}$$

where $\|w\|^2$ is Euclidean norm, which denotes the distance between hyperplane and margins, b is a constant, and $\cdot$ expresses a dot product operation.

For non-separable case, introducing slack variables ξ_i (Vapnik 1995), Eq. (3) can be modified as

$$y_i ((w \cdot x_i) + b) \geq 1 - \xi_i \tag{4}$$

Then, introducing $v(0, 1]$ to account for misclassification (Scholkopf et al. 2000), Eq. (1) can be defined as

$$L = \frac{1}{2}\|w\|^2 - \frac{1}{vN}\sum_{i=1}^{N}\xi_i \tag{5}$$

When using a two-class SVM in landslide susceptibility assessment, it is possible to introduce a dot product in which can be represented by a Kernel function $K(x_i, x_j)$ (Vapnik 1995) to account for nonlinear decision boundary.

Moreover, slope units are used as basic mapping unit for the investigation of their effect on susceptibility mapping. Each slope unit is an aggregated object of contiguous grids and is summarized with a set of ground conditions which are different from its adjacent units. The study area was partitioned into 279 slope units derived semi-automatically from the DEM, and its average, maximum, and minimum area are 0.1629, 1.3366, and 0.01646 km^2, respectively.

5 Selection of Predictor Variables

5.1 Topographic Characteristics

Topographic characteristics include slope, aspect, curvature, distance from drainage, slope unit altitude, catchment area, and catchment height (Fig. 3). Slope and aspect of each grid were calculated from the altitudes of adjacent eight DEM points. Tangential curturve was generated from the DEM data to describe the complexity of terrain. Distance from drainage was created from polyline data of drainage using Euclidean distance method. Slope unit altitude was obtained using zonal statistical analysis of slope units in the DEM layer. Moreover, catchment area and catchment height were derived from the DEM using hydrological analysis methods.

5.2 Geological Information

Geological factors include engineering rock group, distance from faults, and slope structure. Engineering rock group units of the study area were grouped into soft rock and alternate hard and soft rock. The closer the distance is to drainages, the higher the landslide susceptibility will be. Distance from faults was calculated from polyline data of geological faults using Euclidean distance method. Slope structure was identified by analyzing the relationship between formation dip and slope (Fig. 4).

Fig. 3 Seven topographic thematic maps
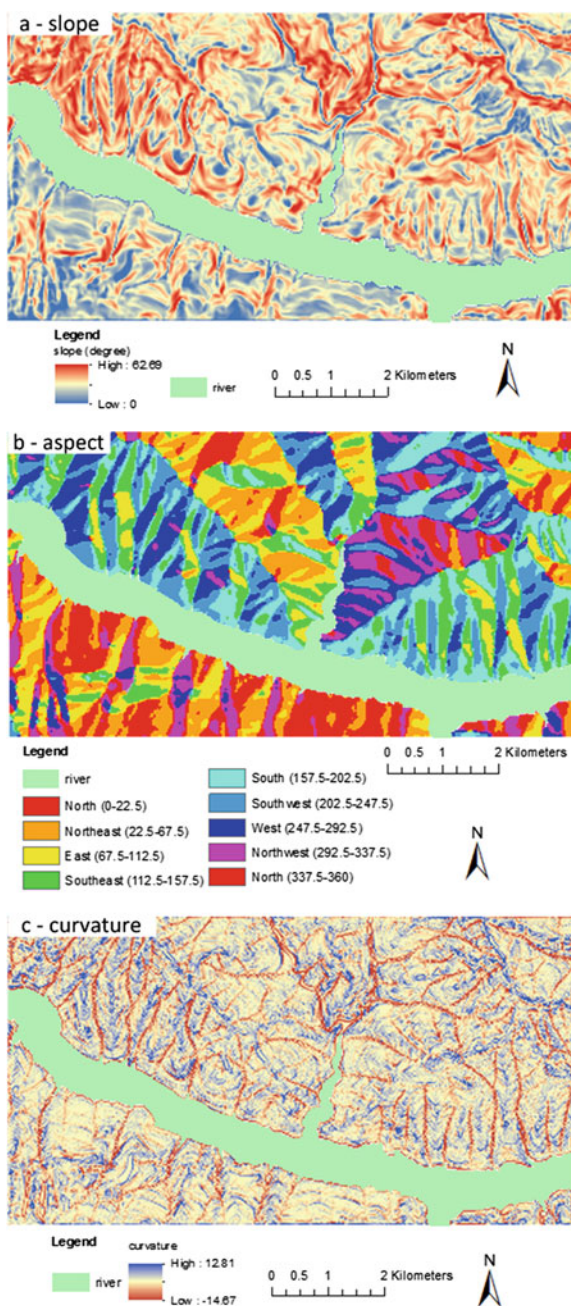
Fig. 3 continued
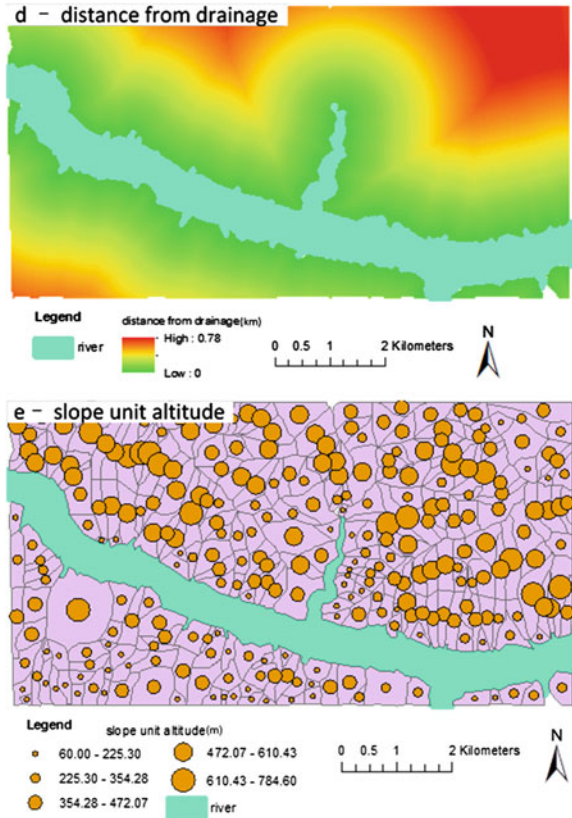
5.3 Land Coverage and Environmental Parameters

Normalized difference vegetation index (NDVI) can quantify the density of vegetation. NDVI was taken into consideration as a landslide-related factor and calculated by Eq. (6)

$$NDVI = (IR - R)/(IR + R) \tag{6}$$

where *IR* denotes the infrared portion of the electromagnetic spectrum, *R* expresses the red portion of the electromagnetic spectrum. Land use information was extracted from the HJ-1A image, and the results were validated by field investigation. In addition, reservoir water level was also a crucial factor for landslide occurrence, provided by Three Gorges Headquarters (Fig. 5).

Fig. 3 continued
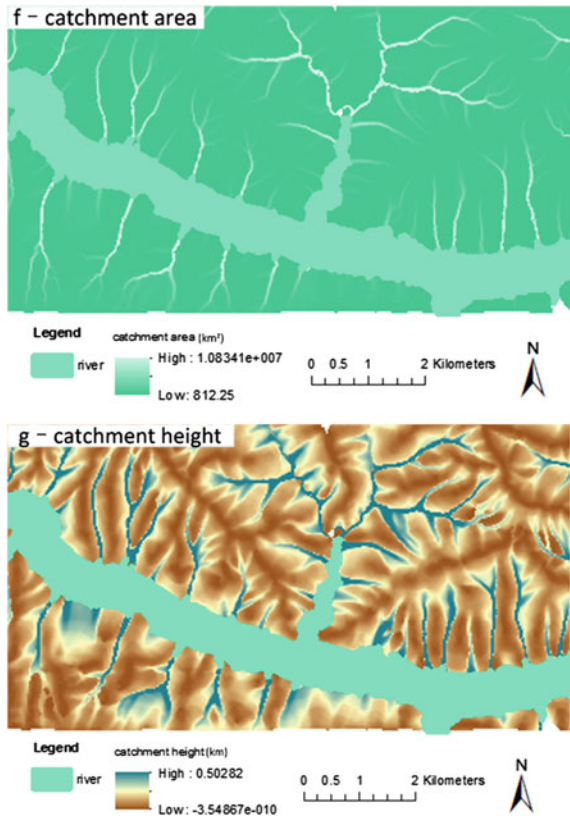
6 Landslide Susceptibility Mapping

For the purpose of landslide susceptibility assessment, the thirteen thematic layers were converted to a 28.5 m × 28.5 m grid. Then slope units were used to resample these environmental factors based on a majority function out of zonal statistics. Thus, dataset of the study area was characterized by 279 rows (number of slope units) and 14 columns (predictor variables and landslide data). The dataset was grouped into 80 % training data and 20 % testing data randomly.

Parameters yielding the highest accuracy were selected as the optimal one. A two-class SVM using the optimal parameters has been trained and then used to calculate landslide susceptibility indices (LSI) of slope units. Radial basis function (RBF) was employed as kernel function (Table 1). The index values of all slope units vary from 0 to 1, which corresponds to landslide susceptibility from low to high. Using such values, landslide susceptibility maps were created. Median of each map is specified as the threshold, the continuous index values can be classified into risk and stable zones. Thus, the landslide susceptibility map can be

Fig. 4 Three geological
thematic maps
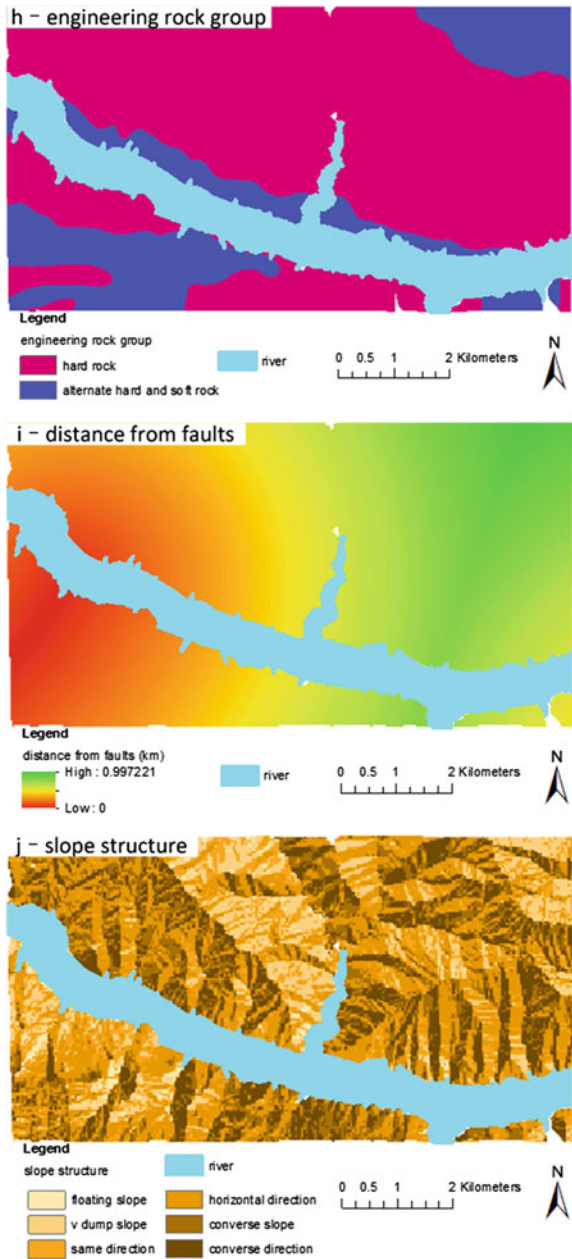
represented by a landslide prediction result map (Fig. 6), continuous LSI map
(Fig. 7), and landslide susceptibility zoning map in which LSI was ranked by
natural breaks algorithm (Fig. 8).

Fig. 5 Thematic maps of NDVI, land use, and reservoir water level
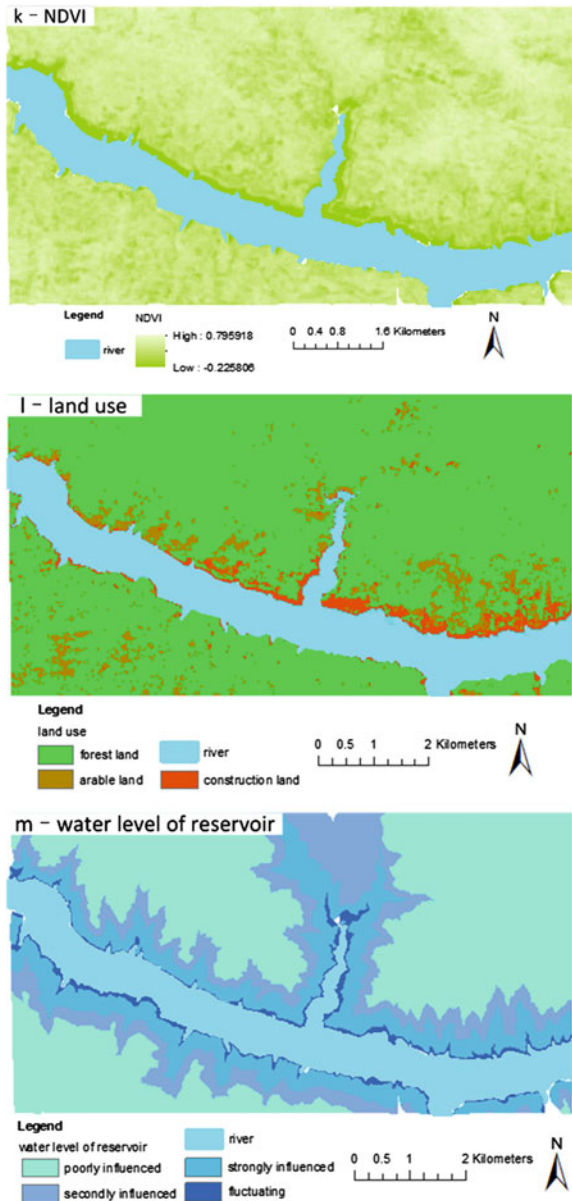
To assess the performance of the two-class SVM, the prediction result was compared with the ones of back propagation neural network (BPNN). The prediction accuracies of the two-class SVM and BPNN were listed in Table 2. Assessment indices consist of overall precision, Kappa coefficient (K), risk area precision, and stable area precision (Wang and Niu 2010). Overall precision is

Table 1 Parameters of the two-class SVM

Parameters	Values
Stopping criteria	1.0E−03
Regularization parameter (C)	10
Regression precision (epsilon)	0.1
Kernel type	RBF
RBF gamma	0.46

Fig. 6 Landslide prediction result map produced by the two-class SVM
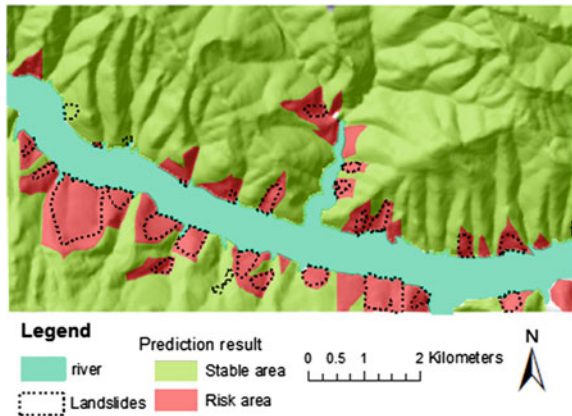
Fig. 7 Continuous LSI map produced by the two-class SVM
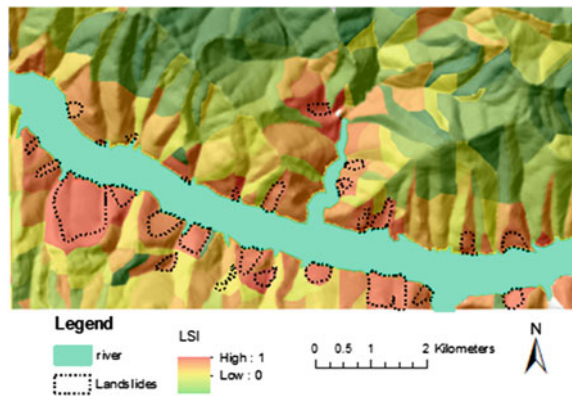
represented by a percentage of slope units number predicted accurately in the whole number of slope units. K is an efficient index to measure association between classification result and actual field investigation. Risk area prediction precision is probability of landslide slope units actually predicted to be landslides. Stable area prediction precision is probability of stable slope units actually predicted to be stable areas. The two-class SVM is obviously superior to the BPNN, and its overall precision, risk area precision, and stable area precision are 93.91, 89.29, and 94.42 %, respectively.

Fig. 8 Landslide susceptibility zoning map produced by the two-class SVM
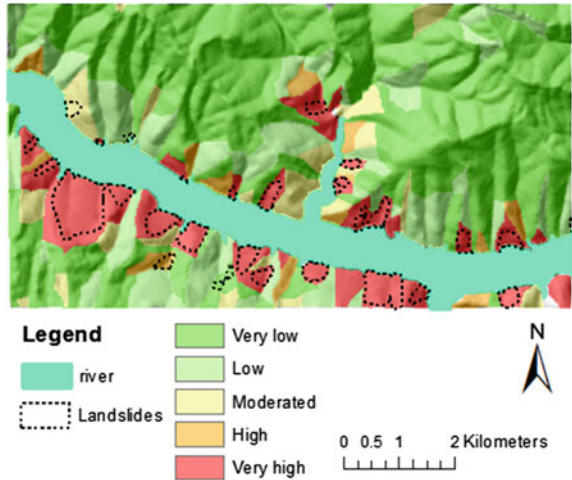
Table 2 Accuracies of the prediction methods

Model	Overall precision (%)	K	Risk area precision (%)	Stable area precision (%)
SVM	93.91	0.5268	89.29	94.42
BPNN	89.25	0.5068	60.71	92.43

7 Validation of the Models

AUC is a good indicator to evaluate the performance of models (Pradhan 2010; He et al. 2012). The validation method was performed by comparing the existing landslide locations with the predicted results. To obtain the relative ranks for each prediction pattern, the calculated LSIs of all slope units were sorted in a descending order, the ordered slope unit values were divided into 100 classes, with

Fig. 9 AUC representing the quality of the SVM used in landslide susceptibility mapping
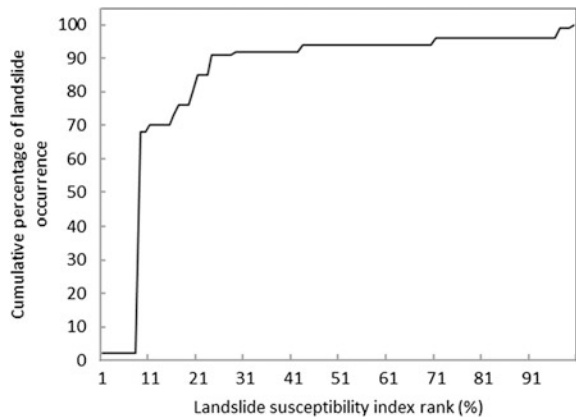
accumulated 1 % intervals. Cumulative percentage of landslide occurrence in classes versus LSI curves was drawn to calculate AUC. The obtained values of AUC from the SVM was given in Fig. 9. The area ratio of SVM were 0.8365, which showed the prediction precision of 83.65 %. The two-class SVM possesses the highest accuracy. For instance, 75–100 % (25 %) class of the study area where the LSI had a higher rank can explicate 91 % of all the landslides using the two-SVM model.

8 Conclusions

This chapter presents a case study based on landslide inventory of the Xietan Town in the Three Gorges, on the application of slope units and a two-class SVM model to produce landslide susceptibility map. To assess performance of the method, the prediction result was compared with other results created by the BPNN according to AUC. From the results, the following conclusions can be drawn:

(1) Multi-source data collection and processing is of great significance to regional landslide prediction. DEM, geological maps, and satellite images serve as the primary data sources and were adopted to obtain important contributing parameters for landslide susceptibility assessment.
(2) According to the formation mechanism of landslides in the Three Gorges, key prediction factors were extracted as input variables of SVM to assess landslide susceptibility.
(3) The slope units-based two-class SVM outperforms BPNN in terms of both accuracy and generalization capacity, with the area ratio 0.8365.

The method proposed in this study can be used as an effective approach for landslide susceptibility assessment in the Three Gorges. It can also be applied to other similar regions by only adjusting the model parameters.

References

Ballabio C, Sterlacchini S (2012) Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. Math Geosci 44:47–70
Bui DT, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree and Naïve Bayes models. Math Probl Eng. doi:10.1155/2012/974638

Chauhan S, Sharma M, Arora MK, Gupta NK (2010) Landslide susceptibility zonation through ratings derived from artificial neural network. Int J Appl Earth Obs 12:340–350

Choi J, Oh HJ, Won JS, Lee S (2010) Validation of an artificial neural network model for landslide susceptibility mapping. Environ Earth Sci 60:473–483

Erener A, Duzgun HSB (2012) Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? Environ Earth Sci 66:859–877

Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Eng Geol 102:85–89

He SW, Pan P, Dai L, Wang HJ, Liu JP (2012) Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. Geomorphology 171–172:30–41

Pradhan B (2010) Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. J Indian Soc Remote Sens 38(2):301–320

Pradhan B (2012) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci. http://dx.doi.org/10.1016/j.cageo

Pradhan B, Lee S (2010) Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. Landslides 7:13–30

Pradhan B, Youssef AM, Varathrajoo R (2010) Approaches for delineating landslide hazard areas using different training sites in an advanced artificial neural network model. Geo-spatial Information Science 13(2):93–102

Rotigliano E, Cappadonia C, Conoscenti C, Costanzo D, Agnesi V (2012) Slope units-based Flow susceptibility model: using validation tests to select controlling factors. Nat Hazards 61:143–153

Scholkopf B, Smola A, Williamson RC, Bartlett PL (2000) New support vector algorithms. Neural Comput 12:1207–1245

Vapnik V (1995) The nature of statistical learning THEORY. Springer, New York

Wang XM, Niu RQ (2010) Landslide intelligent prediction using object-oriented method. Soil Dyn Earthq Eng 30:1478–1486

Yao X, Tham LG, Dai FC (2008) Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. Geomorphology 101:572–582

Yeon YK, Han JG, Ryu KH (2010) Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng Geol 116:274–283

Yilmaz C, Topal T, Lutfi SM (2012) GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak–Turkey). Environ Earth Sci 65:2161–2178

Zarea M, Pourghasemi HR, Vafakhah M, Pradhan B (2012) Landslide susceptibility mapping at Vaz watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab J Geosci. doi:10.1007/s12517-012-0610-x

A New Algorithm for Extracting Drainage Networks from Gridded DEMs

Tao Wang

Abstract Drainage networks are important abstract features in terrain modeling and play functional roles in hydrological, geomorphologic and biological analyzing models. Watershed indices based on drainage networks are crucial in flood predicting models. Huge efforts have been made on automatic extraction of drainage networks. However, there are no effective methods to extract threshold-insensitive and noise-free drainage networks from gridded elevation data. This chapter proposes an algorithm to extract complete and reasonable drainage networks from gridded digital elevation models (DEMs) by integrating global and local methodologies. First, the flow routing algorithm is employed to derive primary drainage segments, which includes depression removal, flow direction computation, flow accumulation and threshold value setting. Threshold values, which are used to filter out meaningful drainage segments from accumulation information, are set based on experts' experience and terrain types. The value heavily influences lengths of individual drainage segments and geometric forms of extracted drainage networks and watersheds, whereas the results are sensitive to threshold values and can introduce uncertainties to further analysis. In order to incorporate the missing drainage segments filtered out by given threshold values, the second step utilizes a moving-kernel method to flag morphometrically characteristic points, which are then integrated into the initial result by downward and upward connecting processes based on flow direction information produced in the first step. Both above mentioned two steps introduce congested drainage segments, which are taken as noises for constructing geometrically clear and topologically consistent drainage networks. In the third step, noisy drainage segments and parallel drainage segments are classified into different types and handled by inductive analysis and rule-based treatment. The topological consistency of drainage networks is maintained in every step. The final results include single-pixel-width drainage networks and correspondent watershed sub-divisions. Two

T. Wang (✉)
Future Cities Laboratory, Singapore-ETH Centre, Singapore, Singapore
e-mail: tawang@ethz.ch

datasets covering different geomorphological areas are used to test our new algorithm. The Strahler ordering scheme, length and structure of extracted drainage networks are analyzed. The quantitative analysis demonstrates that extracted drainage networks based on the new algorithm are insensitive to threshold values. Visual inspection by overlaying extracted drainage networks with contour lines shows that drainage segments are consistent with the contour curvature.

Keywords Complete drainage networks · Digital terrain models · Extracting algorithm

1 Introduction

Drainage information has been playing an important role in modeling geomorphological, hydrological, and ecological processes because of its direct reflection of water convergence and heat redistribution over land surface (Moore et al. 1991; Passalacqua et al. 2010). Digital terrain models (DEMs) are basic data for extracting drainage information. To meet the requirements from research and practice, topographic features have been always taken as a fundamental feature layer on maps and in geospatial databases. Technologies in remote sensing and photogrammetry have been developed for producing high resolution and widely available DEMs, for example the SRTM elevation data (Farr et al. 2007), which further poses new opportunities and challenges in digital terrain modeling. Being a central component of drainage information, drainage networks have been an increasing research focus for both the importance in various spatial models in earth science and the abundant byproducts during extraction process (Mark 1984; Jenson and Domingue 1988; Band 1986; Tarolli et al. 2012; Wilson 2012). In cartography and geographical information science, drainage networks, which are one component of topographic surface networks (Rana 2004), have been also of great interest to facilitate geographic information generalization, data compression and highly efficient terrain analysis algorithms (Wu 1981; Weibel 1992). High density drainage networks are essential prerequisites in many other fields of earth science (Jordan 2007).

The data source for extraction of drainage networks can be in the form of contour, triangulated irregular networks (TINs) or gridded DEMs. The algorithms based on contour lines involve spatial relationship detection among contour lines, curvature computation of contour lines, setting threshold value of characteristic points, linking characteristic points into drainage segments and building of drainage networks. Triangulation or map algebra can be used to facilitate the detection of characteristic points or segments (Aumann et al. 1991; Kweon and Kanade 1994; Ai 2007). However this type of methods suffers from determination of multiple threshold values and ambiguous selection of contiguous points for

connection of segments. In addition, contour lines are generated from gridded DEMs in most cases nowadays, which resamples terrain surface and introduces information loss.

Theoretically, a TIN would be the best data source for extraction of drainage networks because vertices and edges in a TIN are defined as local terrain characteristics. However, an elevation TIN is normally either generated from contour lines or interpolated based on regular elevation grid and most methods for generating a TIN cannot ensure that the edges can meet defined geometric requirements. In order to derive drainage networks, Falcidieno and Spagnuolo (1991) designed Characteristic Region Configuration Graph based on geometric properties between neighboring triangles and defined boundaries lines between two distinctive regions as edges of surface networks. However Kreveld (1997) and Brandli (1996) argued that this algorithm simplifies the complexity of terrain surface and does not work for complex regions. Yu et al. (1996) proposed an algorithm based on a flow routing method that traces steepest downstream lines to generate flow accumulation. Then drainage networks can be filtered out using given threshold values. Zhou et al. (2011) proposed a TIN-based algorithm for drainage information extraction, in which gridded-DEM cells are partitioned into a triangular facet network. Then the flow path of each triangular facet is defined by its aspect and routing to an outlet or a local pit according to downstream neighboring facets. To compute the catchment area for a specific cell, all the upstream cells are summed up by counting flow paths passing this cell. However uncertainties would be inevitably introduced in the transformation from a DEM to a facet TIN. In addition, computations of a DEM to a TIN, of vector flow paths and of catchment area are consuming large computational resources.

Due to a wide availability of gridded elevation data at various resolutions, most research and practical efforts to derive drainage networks are dependent on gridded DEMs. For the same reason, there have been extensive investigations on algorithms of automatic extraction of drainage lines based on gridded DEMs by researchers from the disciplines like GIS, computer science, hydrology and geomorphology. Wood (1996) proposed a classification based on five criteria. Generally there are two types, those based on morphometric analysis methods by template detection or curvature filtering over surface fitting, and those based on simulation of water runoff and flow routing over terrain surfaces.

The morphometric analysis algorithms based on filtering by fixed or variable size kernels are very efficient and easy to be implemented in a parallel environment (Band 1986; Douglas 1986; Bennett and Armstrong 1996). The geometric concavity analyzing can be used to identify characteristic points over terrain surface. It can be implemented in two ways, comparing elevation values on profiles in a given window, or computing curvatures on a fitted polynomial surface. Then successive steps include linking characteristic points into lines, image thinning and further building of drainage networks. These sub-tasks are both difficult to implement and involve multiple threshold values settings. Various methods (Wood 1996; Schneider 2003) have been designed to manipulate them and multi-criterion analysis methods are used, which would introduce uncertainties

into results. Thinning algorithms from image processing are usually used to generate connected drainage lines of a single pixel width, in which physical characteristics of terrain surface obviously are not considered. Another disadvantage of these algorithms is that further processing has to be done to extract watershed subdivisions and other drainage information and to correlate them with drainage networks.

The second type of algorithms has a physically sound theoretical assumption which imitates water runoff flowing and accumulating process over terrain surface (O'Callaghan and Mark 1984; Jenson and Domingue 1988; Freeman 1991; Martz et al. 1992; Planchon and Darboux 2001; Endreny and Wood 2003). The basic assumption is that water runoff drains out of land into the sea along downslope flow paths. Areas where the accumulated amount of flow paths is greater than a given threshold value are taken as drainage channels. In this assumption, every point on land surface should have one or multiple downstream flow directions. However, the points in depressions or pits which are surrounded by higher points and in flat areas cannot find proper flow directions. Those points would block upstream water runoff and affect correct generation of accumulation. There have been various ways to handle depression areas, for example image smoothing to eliminate small depressions (O'Callaghan and Mark 1984), elevating all points to the lowest pour height (Jenson and Domingue 1988; Wang and Liu 2006), leveling higher points along flow paths down to pit's height (Martz and Garbrecht 1998). All of the strategies have influence on the forms of extracted drainage networks and a hybrid approach can produce better results comparing to river network from topographic maps (Poggio and Soille 2012). The second critical task is to assign flow directions to all cells in gridded DEMs. Special treatments should be made to points in flat areas which are produced by depression filling or inherent in original DEMs. Flow directions of these points can be assigned by introducing the direction information from rim of depression during the depression filling step (Jenson and Domingue 1988), or by altering elevation values randomly (Planchon and Darboux 2001) or linearly (Pan et al. 2012) to produce a gentle slope. To distribute water runoff from one point to its downslope neighbors, single-flow direction (SFD) and multiple-flow direction (MFD) algorithms have been proposed and evaluated in various scenarios (Zhou et al. 2011). MFD has a sound basis to simulate the divergent flow over ground. However recent investigations demonstrate that MFD algorithms produce broad drainage segments and patterns which are not well defined to show flow convergence and single flow direction algorithms show thier advantages (Zhou et al. 2011; Poggio and Soille 2012).

Considering the two types of methods with gridded DEMs, there are at least two main issues which have not been addressed carefully. Both morphometric analysis algorithms and flow routing algorithms produce broad areal shaped drainage channels, which cannot clearly represent channel positions and further affect parameter calibration in distributed hydrological modeling. Algorithms based on image thinning and operators of mathematical morphology can be used to generate single-pixel-width drainage segments. But apparently physical terrain characteristics are not considered. The second problem is that the methodology using

threshold values can always miss drainage segments partly or even in total, which may in many circumstances fail to identify typical drainage segments, especially in upstream areas. Smaller threshold values produce high density drainage networks. But numerous pseudo and parallel drainage segments normally appear at downstream area (Tarboton 1997). It leads to difficulty of finding correct positions of valley heads and drainage density which are important to investigate geomorphological activities (Tribe 1991).

This chapter describes a new method by employing water runoff flow routing algorithm to derive the global structure of drainage networks, and the morphometric algorithm to supplement missing drainage cells at upstream area. A noise cleaning procedure is designed to remove problematic segments based on hydrological and geomorphological properties and to improve topological consistency of drainage networks. Based on test of two datasets of different terrain types, a quantitative analysis is made to the extracted drainage networks. Conclusions and discussion are provided at the end.

2 Methodology and Algorithms

Algorithms based on water runoff flow routing have a sound physical basis and can generate topologically explicit drainage networks and watershed subdivision by threshold values of accumulation. Morphometric algorithms can produce good results in the areas where drainage segments are well defined. These two types of algorithms are complementary if sensitivity to threshold values can be reduced and a better method can be designed different from image thinning.

2.1 Primary Drainage Networks and Watershed Information

The algorithm of extracting drainage lines based on water runoff routing executes typically five steps. First, all depressions in gridded elevation matrix are to be identified and processed. There are two ways to remove the depressions. The most used one is depression filling, which raises all grid cells in a depression to a new height value equal to the elevation of lowest pour points on the rim of the depression. Another way is dam breaching (Poggio and Soille 2012). In this chapter, the former one is used to ensure that all runoff flow paths can be directed to the boundary of depression regions.

After removal of depressions, every cell is given a flow direction based on deterministic-8 method (D8). The flow direction of a cell points to one neighbor cell which has the steepest downward slope among its eight neighbors. The flat area cells, which are produced by depressions removal or inherent in original data, must be specially treated. The method which was developed by Martz and Garbrecht (1998) is used here to assign flow directions to cells located in flat areas. A

cell is a neighboring upstream cell to the one that its flow direction points to. A cell is a neighboring downstream cell to the one that is pointed by its flow direction. Those cells without upstream (or downstream) cells can be starts (or ends) of flow paths.

Each cell is assigned one unit of water as runoff on this point. Accumulation process is started from all starts of flow paths and arrives at boundary of data or research region along flow paths. Accumulation value of one cell is the number of its upstream cells and is taken as its drainage area. The results include a flow direction matrix and a drainage area matrix, as showed in Fig. 1b. The process corresponds to the physical process of water runoff over terrain surface.

A threshold value is given in the fourth step and the drainage area matrix is segmented into a binary matrix. The grid cells with higher drainage area are marked as drainage cells and constitute drainage segments connected along flow directions. However, as stated above, an appropriate threshold would best be determined according to terrain characteristics and expert experiences. A higher threshold value commonly produces lower density of resulted drainage lines.

The last step of this stage builds the drainage networks, acquires basic information of each drainage segment. The basic information of each drainage segment consists of unique identifier, length, slope, and topological information of connected segments. The topological connectivity combined with flowing directions can be used to assign stream ordering numbers to drainage segments to distinguish their hierarchical levels. This chapter uses Strahler ordering system for its popularity in hydrological applications. One drainage segment with no other segments draining in is marked as first order (as light pink segmetns in Fig. 1c). Two or more segments with equal order can join together and form a segment with one-level higher order. A segement uses the higher order number if two segments with different order numbers join together. Then all cells along one flow path are assigned a sub-watershed number according to the drainage segment that the flow path drains into.

Each drainage segment and its associated sub-watershed share one unique identifier which is started at 1. 0 denotes cells out of research region, drains out of data boundary without enough accumulation or simply no data. The sub-watershed identifiers and Strahler orders are stored in matrices and can be converted to vector structures when needed. The implementation is showed in Fig. 2. In the figure, *StrhSta* is an integer array and stores the number of drainage segments on each Strahler order (SID). For each grid cell P, the values of S_P, O_P and I_P denote corresponding Strahler order, unique identifier and flowing-in number. OID denotes the object identifer number of each drainage segment and the corresponding watershed. The topological information of drainage segments is recorded during the iteration. The length and slope can be computed based on original elevation, Strahler order and identifier matrices.
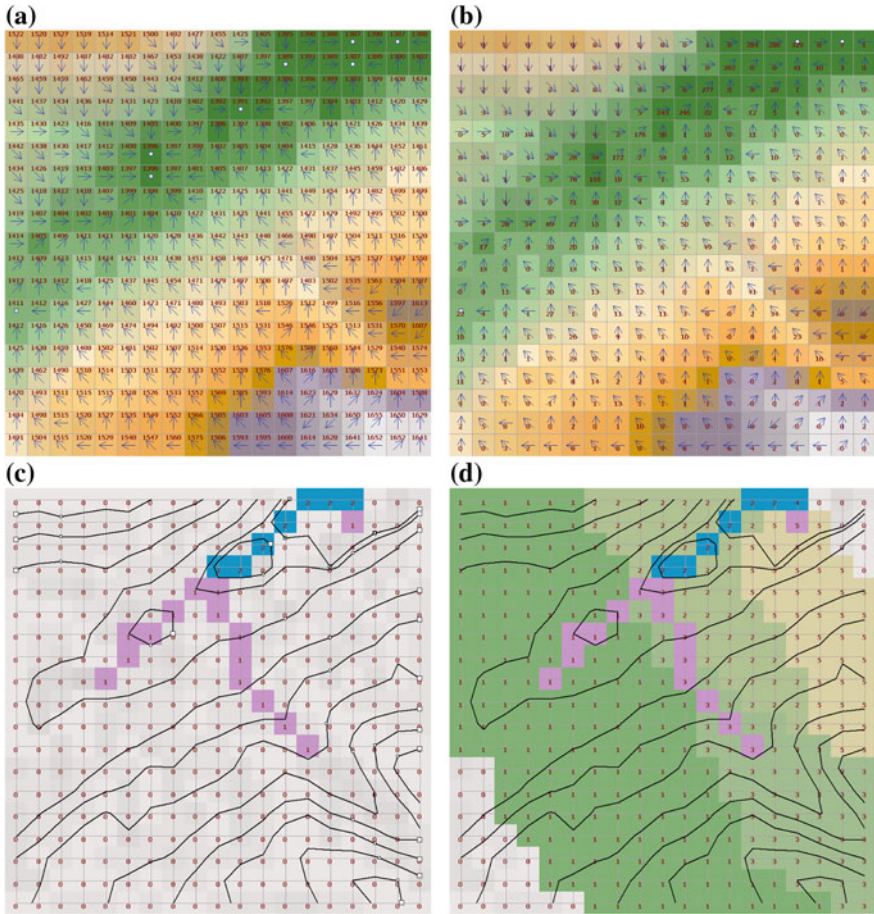
Fig. 1 Flow directions of grid cells with elevation value; **a** Flow directions over original elevation matrix. Each value denotes one cell's elevation. **b** Accumulation matrix with flow directions. Each value denotes amount of upstream runoff. **c** Drainage segments filtered out by threshold value 40. Values greater than 0 denote Strahler orders. 1 and 2 denote first and second order in *light pink* and *blue*. **d** Watershed subdivisions. Values greater than 0 denote watershed numbers. The two closed contour lines are located around the two depressions.

2.2 Complementary Drainage Lines

The threshold value in above processing dramatically influences drainage density. Figure 3 shows the changing patterns of identified drainage segments with two threshold values. The value is normally set to as small as possible in order to identify high density drainage segments. However it is difficult to set an appropriate one.
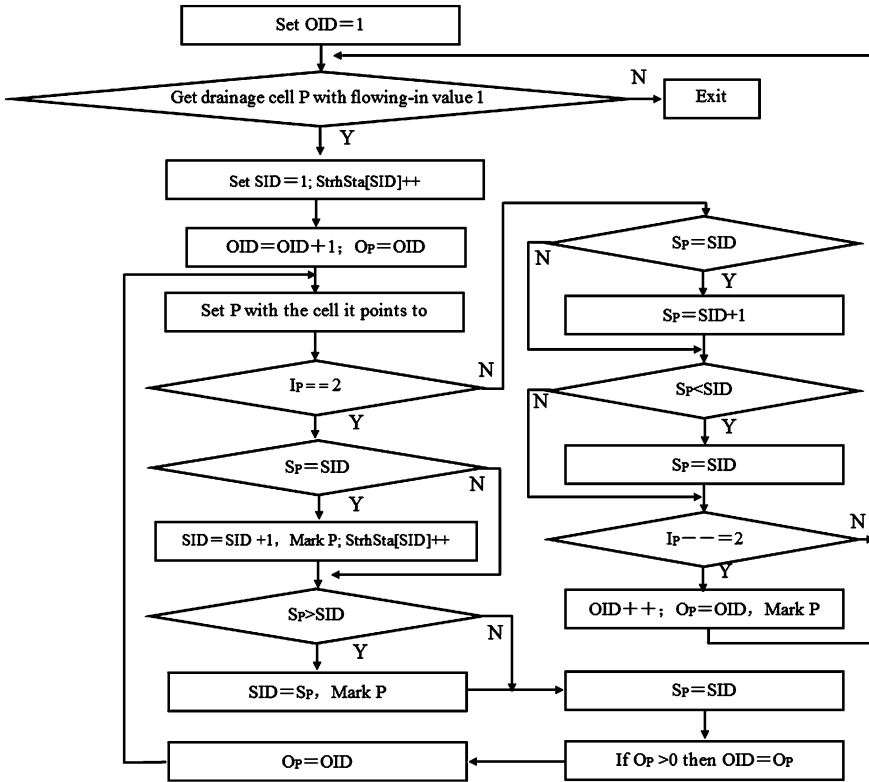
Fig. 2 Extraction of drainage information

From visual inspection of the map containing contour lines and identified drainage segments, we can see that most missing drainage segments are actually located at upslope areas, where water runoff accumulation may not be significant but the geometric shapes of valleys are obvious. At this kind of situation, morphometric analysis has its advantage.

The second stage of our algorithm, which includes concavity analysis in bottom-up and top-down strategy, complements the initial results into a complete version. All the local terrain characteristic cells are identified using a moving 3 by 3 kernel (Douglas 1986). Valley cells are marked as "VL". The channel head identified in primary network detection step is used in a bottom-up process. Then all the start cells of first order Strahler drainage segments are used to connect neighboring valley cells which drain into start cells. Figure 4a shows the growing result of Fig. 3a. The top-down step is executed based on the local terrain characteristic cells with "VL" which have not been identified as drainage cells but marked as valley cells. These cells flow along the steepest downward slope until reaching current drainage cells which are not necessarily first order drainage segments. In order to acquire reasonable results, the two steps consider segments'
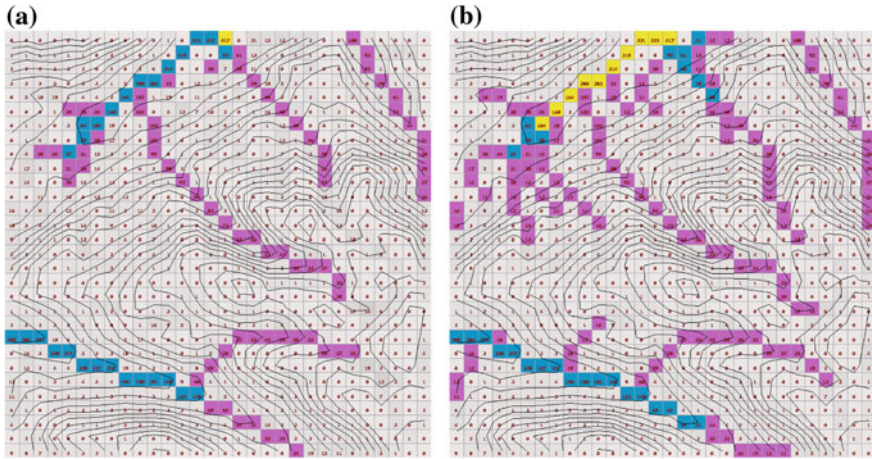
Fig. 3 Changes of drainage patterns with different threshold values. *Yellow cells* are third Strahler order. **a** Threshold value is 20. **b** Threshold value is 10

slope during complementing process. The characteristic cells on flat area cannot be connected into drainage lines. The drainage lines are then organized into topological networks and labeled with Strahler order.

3 Rule-Based Treatment of Problematic Segments

Drainage networks extracted above are organized into a topological structure during Strahler ordering procedure. However, when a threshold value is set to small and after the complementary step in Sect. 2, a large amount of problematic drainage segments may be generated in various geometric forms. Parallel segments over flat area are often produced. Congested segments are those that partly or all of cells are direct neighbors of cells on other drainage segments and they do not flow into each other. Here only cardinal neighborhood between cells is considered (4-connectivity). This situation appears everywhere, as showed in Figs. 3 and 4. In general it makes areal shaped drainage segments and fragmented watershed sub-divisions. Image thinning methods used in previous research are apparently not appropriate because all cells are treated indiscriminately disregarding flow directions. In order to remove problematic cells, drainage segments are first classified into different categories based on types of neighboring cells, flow directions, Strahler ordering numbers and characteristic significance (valley, saddle or ridge point). Special treatment methods are designed to handle different situations.
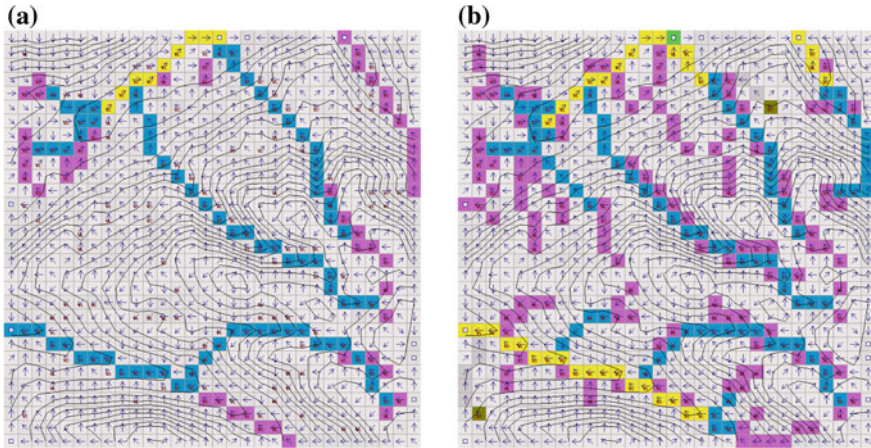
Fig. 4 Complementary drainage segments based on geometric characteristic analysis. VL in the cells denote grid cells of valley; **a** bottom-up growing; **b** top-down connection

3.1 Straight Drainage Segments

This kind of drainage segments mostly appear as first Strahler order located in lower elevation areas. The shape of a typical valley is nearly straight on 2D and the slope is small on terrain profile. The drainage accumulation value along its flow path does not change remarkably because there are few other segments draining in. The cells on this kind of segments are normally not "VL" cells marked in morphometric detection. The corresponding cells are simply removed from the drainage segments (as dark green in Fig. 5). The Strahler orders of downstream drainage segments are updated after this process to maintain the topological structure of drainage networks.

3.2 Congestion of First Order Drainage Segments

After the complementary cells marked according to morphometric characteristics are appended, a large number of drainage segments are congested. In order to maintain topological connectivity of drainage networks, the cleaning process is initiated from the segments with first Strahler order. Three types of congestion are further categorized for corresponding treatments.

(1) One-cell length segments. If the only cell of a 1st-order segment has more than one cardinal neighboring cell with higher order, or more than two cardinal neighboring cells with 1st order which are not at downstream direction of it, it is removed directly (the dark gray cells in Fig. 6b).
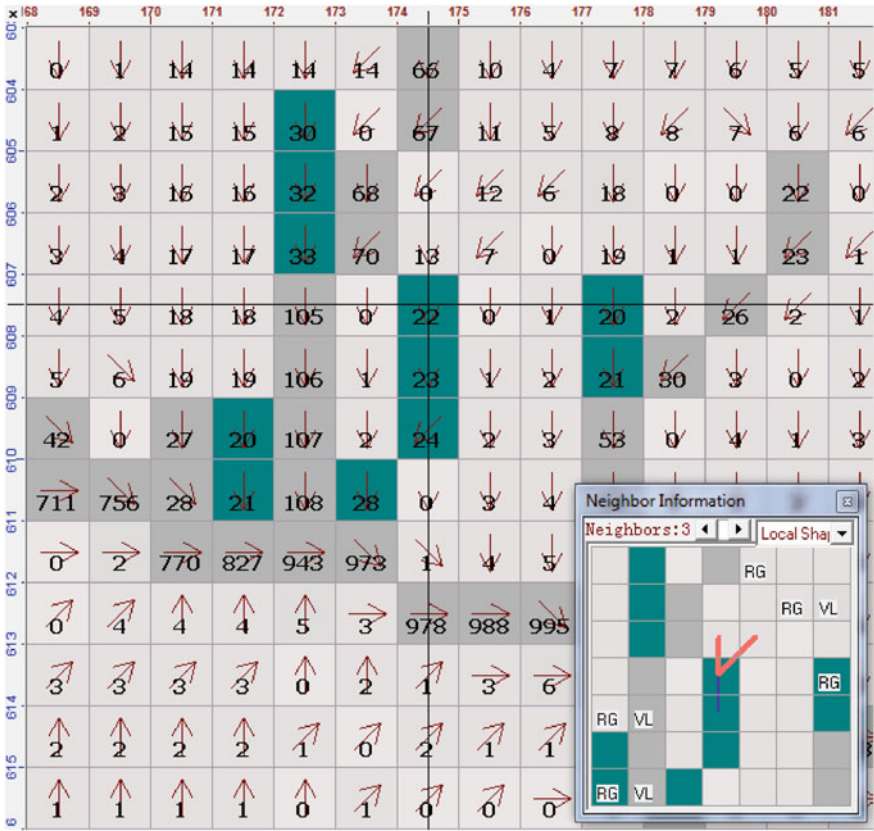
Fig. 5 Removal of *dark green* straight drainage segments. Cells in *light gray* are remaining drainage segments. Values in each cell are drainage area value. VL in the small window denotes valley cells

(2) Congestion between 1st-order segments. If there exists two adjacent segments of this type, the one with the higher accumulation value, longer length, larger slope and smaller flow direction change is to be preserved. These criteria are applied consecutively until one is met. If both segments show identical configuration considering all criteria in very rare conditions, the first one is to be removed. If the removed segment is partly left, the cell located furthest downstream need to find a new neighboring cell as its pour point to ensure topological consistency of drainage networks. The new downstream cell is selected according to lowest elevation and highest accumulation value. The accumulation values of all cells on the two affected flow paths are updated afterward. If more than two 1st-order segments are congested together, the middle one is to be reserved to keep the final result balanced geometrically. The congestion involved more than 3 segments is very rare.
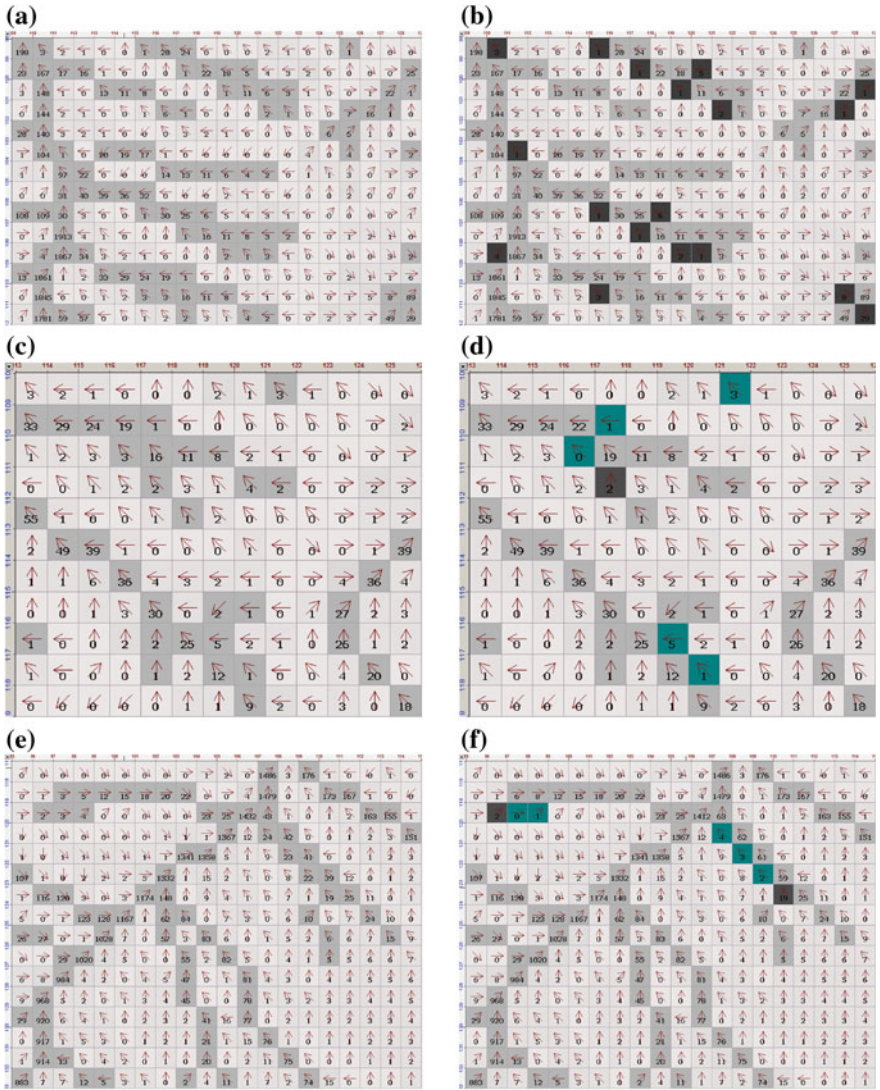
Fig. 6 Removal of congested drainage segments at first Strahler order; **a** Drainage cells before processing (1). **b** Drainage cells after processing (1). **c** Drainage cells before processing (2). **d** Drainage cells after processing (2). **e** Drainage cells before processing (3). **f** Drainage cells after processing (3)

As showed in Fig. 6c, d. The dark green cells are removed due to congestion between 1st-order segments. The one which is partly removed has to change the flow direction of its last downstream cell (in dark gray) to appropriate neighboring cells.

(3) Congestion of 1st-order segments with higher-order segments. A 1st-order drainage segment can be congested with higher order segments in part or in whole. The congested cells on 1st-order drainage segments are to be removed and higher order neighbor cells leave unchanged. If the middle part of one 1st-order order segment is congested with other segments, the congested part is removed and this segment is broken into two segments.

The flow direction of last downstream cell of remaining segment is changed and the accumulation values along old and new flow paths are to be updated in the same way as above.

3.3 Congestion of Higher Order Drainage Segments

This section deals with congestion occuring between segments with Strahler order equal to or larger than 2. When there are congested cells between two segments with different Strahler orders, the lower one is to be removed. If the Strahler orders are equal, the one with smaller drainage area, short length is to be processed. The processing starts from the downstream outlet of the target segment. The congested cells are removed until the last congested cell. Local flow directions and drainage area on flow path are altered at the same time with similar strategy in above section. Figure 7 shows a 2nd-order segment (in blue) is partly modified to avoid congestion with a 5th-order segment (in pink).
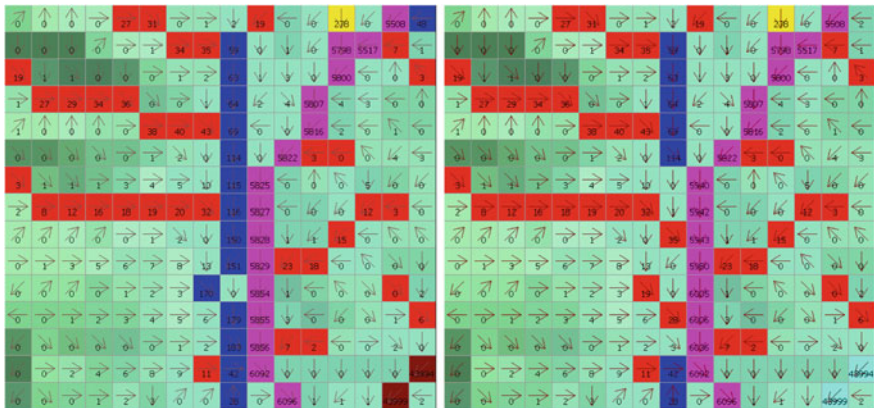
Fig. 7 Lower order segment is partly removed to reduce congestion with neighboring higher order segment

3.4 Processing Iteration

The above situations account for most congestion. However, the processing must be iterated for multiple passes because the removal of higher Strahler order congestion may produce a new situation. As in Fig. 7, after congestion removal, there are two 1st-order segments generated, which may generate further congestion.

4 Evaluation and Results

The new algorithm is tested with two DEM datasets from SRTM version 2.1 (SRTM 2012), whose original resolution is 3-arc seconds. The first one is the data tile named N36E110 with 1201 rows by 1201 columns. It covers an area in China with of loess landform shaped by erosion processes of wind and water runoff, where the Yellow River runs through. The elevation ranges from 414 to 1803 m. The second data is located at upstream area of Ciliwung River, West Java of Indonesia with 1000 rows by 1200 columns. It covers several eroded volcanic mountains. The elevation ranges from 5 to 3006 m. Figure 8 shows the terrain shading maps of the two testing datasets.

In order to perform an evaluation of the results of the new algorithm, the flow routing algorithm and the new algorithm are implemented and applied with different accumulation threshold values (T) on the two testing datasets. The extracted drainage segments are organized into networks labeled with Strahler orders using the procedure showed in Fig. 2. The amount of drainage segments at different Strahler orders is listed in Tables 1 and 2. The rows in gray background are results from the new algorithm.
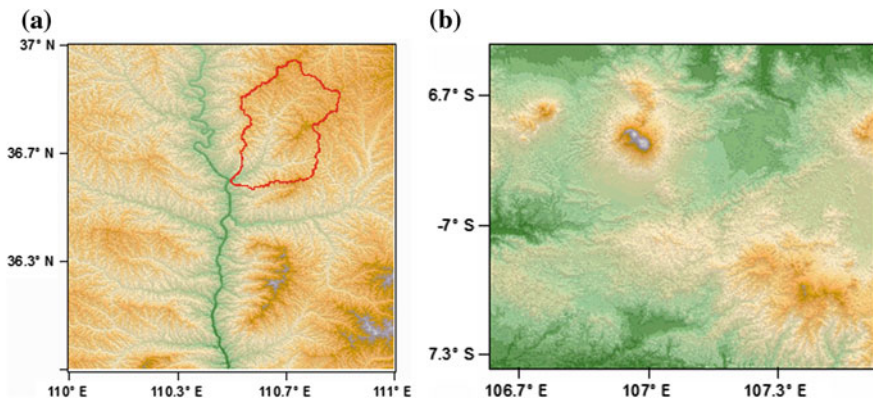
Fig. 8 Two test DEM datasets. The sub-watershed with the *red* boundary in test dataset 1 is to be highlighted; **a** Test dataset 1; **b** Test dataset 2

Table 1 Result of drainage networks extracted from dataset 1. T denotes the threshold value for accumulation filtering

T	Strahler order								Sum of	
	1	2	3	4	5	6	7	8	Segments	Length
100	3778	781	152	30	8	1			4750	85553
50	7243	1446	296	66	13	3	1		9068	116916
100	68817	14085	3179	690	141	32	8	1	86953	388609
50	68819	14086	3172	691	141	32	8	1	86950	388656

Table 2 Result of drainage networks extracted from dataset 1. T denotes the threshold value for accumulation filtering

T	Strahler Order								Sum of	
	1	2	3	4	5	6	7	8	Segments	Length
100	3067	713	170	43	10	3			4006	82535
50	5783	1298	325	86	17	4	1		7514	110926
100	48938	10158	2369	557	137	36	6	2	62203	309060
50	48944	10172	2362	564	141	36	6	2	62227	309302

The statistics of extracted drainage networks in Tables 1 and 2 shows that threshold values have heavy influence on results of flow routing algorithm. The total number and the length of drainage segments, the Strahler ordering levels and the numbers of segments on each level are changed remarkably when threshold values are different. But the results from our algorithm are very consistent and stable concerning all indicators. The fluctuation at lower Strahler orders is due to the fact, that lower threshold values identify non-valley cells at low and flat valley bottom.

A sub-watershed in the first dataset (red boundary in Fig. 8a) is extracted out for a more careful comparison. Figure 9 shows results of the flow routing algorithm and the new algorithm. Figure 9d shows drainage segments are identified at every contour bends. The numbers of drainage segments at all Strahler orders are listed in Table 3. The corresponding bifurcation ratios are computed. It is apparent that the accumulation threshold values have large influence to this indicator in the flow routing algorithm. The ratio based on the new algorithm is stable to the changes of the threshold values.
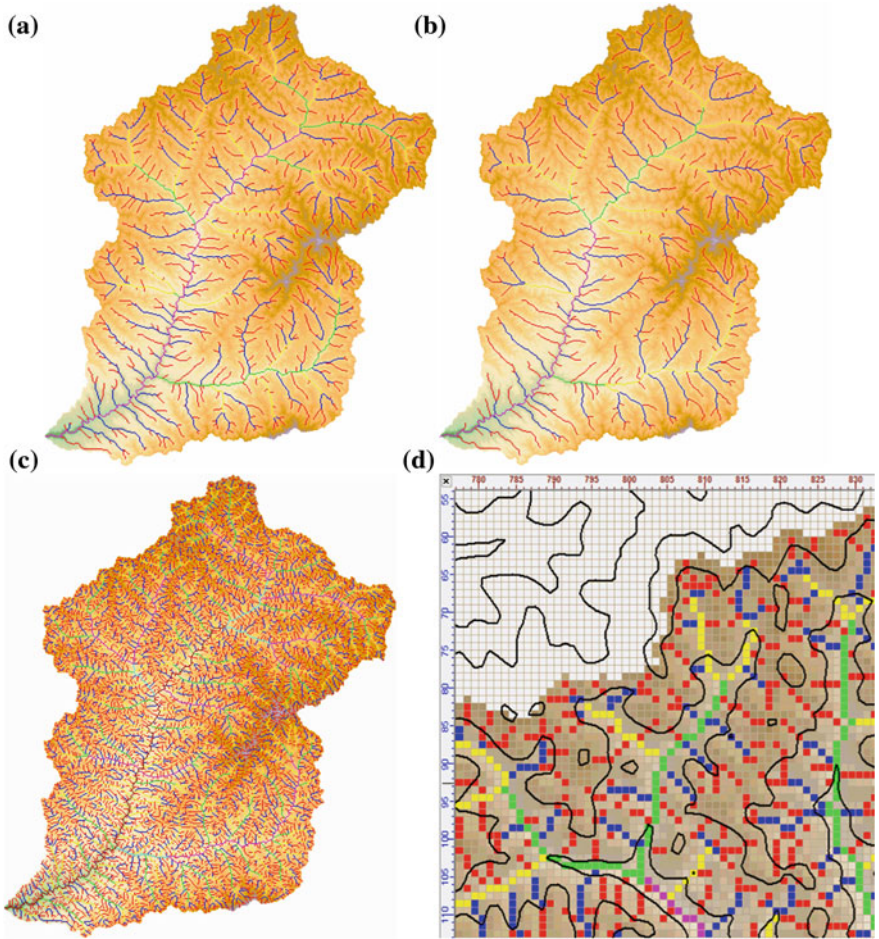
Fig. 9 The extracted drainage networks from ordinary algorithm and the new one. *Red*, Pixels in *blue*, *yellow*, *green* and *pink* are in 1st, 2nd, 3rd and 4th orders. *Solid black curves* are contour lines. **a** Drainage network at T=50. **b** Drainage network at T=100. **c** Drainage network of new algorithm. **d** Overlay of drainage with contour lines

Table 3 Result of drainage networks of a sub-watershed

T	Strahler order								Bifurcation ratio
	1	2	3	4	5	6	7	8	
100	304	69	11	3	1				3.47
50	583	121	24	6	1				3.97
100	6011	1253	293	62	12	4	1		3.71
50	6011	1253	291	62	12	4	1		3.71

5 Conclusion and Discussion

This chapter reviews the existing algorithms of drainage network extraction and finds that there are no eligible algorithms, which can produce complete and reasonable drainage networks from gridded DEMs. Based on analysis of inherent rationale of available research, we proposed and implemented a new method that integrates flow routing algorithm and revised local morphometric algorithm. A rule-based cleaning process is designed to evaluate and eliminate problematic and noise segments. Two example datasets of different terrain types were used to test the proposed method. The quantitative analysis of extracted drainage networks shows that results by this new algorithm are insensitive to threshold values and stable in hierarchical structures.

Further possible development based our work can be on different aspects. Multiple-flow-directions algorithms are more appropriate to simulate water diverging models in microtopography, especially as high resolution DEMs are more and more available from wide applications of LiDAR and low altitude aerial stereo images. Also in this scenario, as natural sinks or depressions are widespread, adaptable pit removal algorithm should be implemented by integration of filling and carving based on identifying objects that break flow routing over terrain surface. In low and flat downstream areas, there are many efforts to design reasonable treatment algorithms to generate flow routing. An alternative and more appropriate way is to use hydrological features and water bodies from large scale topographic maps and geospatial databases as an a priori knowledge, which are then incorporated to reduce problematic segments during or subsequent to drainage networks extraction process. At the same time, a geospatial database can be enriched by incorporating the extracted drainage information (flow direction, flow accumulation and watershed subdivision). Another work is to design an appropriate generalization method to simplify the very dense drainage networks of the new algorithm for various applications.

References

Ai T (2007) The drainage network extraction from contour lines for contour line generalization. ISPRS J Photogram Remote Sens 62(2):93–103

Aumann G, Ebner H, Tang L (1991) Automatic derivation of skeleton lines from digitized contours. ISPRS J Photogram Remote Sens 46(5):259–268

Band L (1986) Topographic partition of watersheds with digital elevation models. Water Resources Research 22(1):15–24

Bennett D, Armstrong M (1996) An inductive knowledge based approach to terrain feature extraction. Cartogr Geographic Inform Syst 23(1):3–19

Brandli M (1996) Hierarchical models for the definition and extraction of terrain features. Geographic objects with indeterminate boundaries. Taylor & Francis, London, pp 257–270

Douglas D (1986) Experiments to locate ridges and channels to create a new type of digital elevation model. Cartographica 23(1):29–61

Endreny T, Wood E (2003) Maximizing spatial congruence of observed and DEM-delineated overland flow networks. Int J GIS 17(7):699–713

Falcidieno B, Spagnuolo M (1991) A new method for the characterization of topographic surfaces. Int J GIS 5(4):397–412

Farr T, Rosen P, Caro E et al (2007) The shuttle radar topography mission. Rev Geophys 45:2005RG000183

Freeman T (1991) Calculating catchment area with divergent flow based on a regular grid. Computers and Geosciences 17(3):413–422

Jenson S, Domingue J (1988) Extracting topographic structure from digital elevation data for geographic information system analysis. Photogramm Eng Remote Sens 54(11):1593–1600

Jordan G (2007) Adaptive smoothing of valleys in DEMs using TIN interpolation from ridgeline elevations: an application to morphotectonic aspect analysis. Comput Geosci 33(4):573–585

Kweon I, Kanade T (1994) Extracting topographic terrain features from elevation maps. J Comput Vis, Graph Image Process: Image Underst 59(2):171–182

Mark D (1984) Automated detection of drainage networks from digital elevation models. Cartographica 21(2):168–178

Martz L, Garbrecht J (1998) The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models. Hydrol Process 12:843–855

Moore I, Grayson R, Ladson A (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5(1):3–30

O'Callaghan J, Mark D (1984) The extraction of drainage networks from digital elevation data. Comput Vis Graph Image Process 28(4):323–344

Passalacqua P, Do Trung T, Foufoula-Georgiou E, Sapiro G, Dietrich W (2010) A geometric framework for channel network extraction from lidar: nonlinear diffusion and geodesic paths. J Geophys Res 115:F01002

Planchon O, Darboux F (2001) A fast, simple and versatile algorithm to fill the depressions of digital elevation models. In: Auzet V, Poesen J, Valentin C (eds) Catena Special Issue, Elsevier 46(2/3):159–176

Pan F, Stieglitz M, McKane R (2012) An algorithm for treating flat areas and depressions in digital elevation models using linear interpolation. Water Resourc Res 48:W00L10

Poggio L, Soille P (2012) Influence of pit removal methods on river network position. Hydrol Process 26:1984–1990

Qian J, Ehrich R, and Campbell J (1990) DNESYS-an expert system for automatic extraction of drainage networks from digital elevation data, IEEE Transactions on Geoscience and Remote Sensing 28(1):29–45

Tarolli P, Sofia G, Fontana G (2012) Geomorphic features extraction from high-resolution topography: landslide crowns and bank erosion. Nat Hazards 61:65–83

Rana S (2004) Topological data structures for surfaces: an introduction to geographical information science. Wiley, Chichester

Schneider B (2003) Surface networks: extension of the topology and extraction from bilinear surface patches. In: Proceedings of GeoComputation 2003, University of Southampton, UK, CD-ROM

SRTM data. http://www2.jpl.nasa.gov/srtm/. Accessed 1 Oct 2012

Tarboton D (1997) A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour Res 33(2):309–319

Tribe A (1991) Automated recognition of valley heads digital from elevation data. Earth Surf Proc Landf 16(1):33–49

Van Kreveld M (1997) Digital elevation models and TIN algorithms. Algorithmic foundations of geographic information systems. Springer, Berlin, pp 37–78

Wang L, Liu H (2006) An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. Int J GIS 20(2):193–213

Weibel R (1992) Models and experiments for adaptive computer-assisted terrain generalization. Cartogr Geogr Inf Syst 19(3):133–153

Wilson J (2012) Digital terrain modeling. Geomorphology 137:107–121

Wood J (1996) The geomorphological characterisation of digital elevation models. City University PhD Thesis

Wu HH (1981) Prinzip und Methode der automatischen Generalisierung der Reliefformen. Nachrichten aus dem Karten- und Vermessungswesen Heft Nr. 85:163–174

Yu S, van Kreveld M and Snoeyink J (1996) Drainage queries in TINs: from local to global and back again. In: Proceedings of 7th international symposium on spatial data handling, pp 13A1–13A14

Zhou Q, Pilesjö P, Chen Y (2011) Estimating surface flow paths on a digital elevation model using a triangular facet network. Water Resour Res 47:W07522

Part VII
User Generated Content and Spatial Data Infrastructure

The Visitors: A Collective Methodology for Encountering and Documenting an Unfamiliar Cityscape

Laurene Vaughan

Abstract From the Dadaist and Situationist walking interventions, to contemporary locative media events and gaming, multimodal mapping methods have been used to extend vernacular methods for knowing and experiencing place, typically cities. In June 2012 a collective of artists, designers and filmmakers converged at ETH Zurich to participate in the *Cartography and Narrative* Workshop: we named ourselves *The Visitors*. This was an interdisciplinary initiative of the Art and Cartography Commission of the International Cartographic Association. Within the context of the Workshop activities, this collective were drawn together out of a shared interest in the 'unknowness' of unfamiliar cities, and a desire to map place through encounter; mapping through walking in particular. The outcome of this collective 'derive' is a film—*The Visitors*—a time based digital map, that embraces the ambiguities of subjective mapping and place-making.

Keywords Imprecise geography · Collaboration · Film

1 Introduction

Cartography (map-making) is 'the discipline dealing with the conception, production, dissemination and study of maps in all forms' (International Cartographic Association 1995). Another description of cartography is that it is the 'Art, Science and Technology' of map making. This is an important definition in that it leads us to appreciate that cartography covers many disciplines and is so wide-ranging that it invites interest from a huge diversity of people. This results in a subject, be it

L. Vaughan (✉)
School of Media and Communication, RMIT University, Melbourne, Australia
e-mail: laurene.vaughan@rmit.edu.au

commercial or academic, that includes people from all backgrounds and specialisms, making a career in cartography both diverse and exciting.[1]

As noted by Cacquard (2011) there is a growing body of scholarship and practice exploring disciplinary differences and intersections (in both the practices and resultant artefacts) in relation to mapping, cartography and narrative. When considered within the framework of The British Cartographic Society definition of cartography, this phenomenon seems unsurprising, and yet this is a rich space of innovation. There is an increasing number of explorations into the form, agency and integration of different mapping methods, including technologies that enable and embed geospatial representation across the realm of contemporary communication contexts.

Over the past five years a global community of inquiry within the Art and Cartography Commission of the International Cartographic Association, has been exploring these intersections through a range of publications, workshops and most recently in the co-production of a film. One such event was the Cartography and Narratives workshop which was held June 11–13, 2012 at ETH Zurich. This event was convened by Barbara Piatti and Sebastien Caquard. It brought together a community of 29 artists, designers, filmmakers, cartographers, geographers and cultural theorists, all with an interest in the rich potential at the exists at the intersection of cartography and narrative. Abstract submissions from the participants revealed that many people were interested in maps and cartography not in terms of geospatial placement and certainty,[2] but rather as a poetic, humanistic or cultural phenomenon. Their proposals, and the subsequent workshop outcomes, celebrated the ways in which artistic and personal narratives create imprecise geographies (Piatti et al. 2009). These are what we can call the affective qualities of place-making and place representation.

2 The Project

From the initial call for proposals it was evident that a key ambition of the Narrative and Cartography Workshop was for the participants to collaborate and explore their respective topics of interest, together. The Workshop was attended by an international and interdisciplinary selection of people. This included cartographers, literary theorists, geographers, designers, cultural theorists and artists. Participants were taken through a series of activities that both allowed them to express their particular projects and eventually to collaborate on the co-production of narrative cartographies. On the first day of the workshop each participant presented their individual interest in narrative cartographies to all the workshop participants. Each presentation was preceded by a 30 s movie made by one of the

---

[1] The British Cartographic Society, http://www.cartography.org.uk/default.asp?contentID=720.

[2] http://cartonarratives.wordpress.com

participating artists. These mini movies were the artist's interpretation of the participant project abstract that had been submitted as part of the selection process for participation in the workshop. Following the individual presentations on the first day, participants then self-divided into groups, which were to be led by the artists who had made the introductory movies. This group would spend the rest of the workshop collaborating on an exploration of narrative cartography.

A number of the workshop participants were exploring walking as an area of mapping or cartographic exploration. From walking as a means for reflective inquiry (Vaughan 2012; Gill 2012) to the ramblings of Dadaist inspired derives (Watson 2012) including the integration with social networking media (Bissen 2012), and drawing using GIS technology (Wood 2012)—there was a collective interest in embodied place-making and place-marking. Harnessing the possibility for shared exploration, three of the walking participants elected to work as a group with the artists Schmid and Cramer (Table 1). Initial conversations within the group revealed that we were all interested in the possibilities of working on a collaborative map through a methodology of walking, that would allow each of our approaches to working with space and place, and image and documentation to be present. How this could happen we were unsure. We had no particular place to map, we had five different individual approaches to mapping and recording the experience of place, and had one day to realize an outcome. The group disbanded for the day to think about what could or might be.

On reconvening the following day, the group decided to undertake the making of a collaborative map in some part of Zurich. None of the group members were what might called 'locals' (Lippard 1997) to the city of Zurich, and for three of us this was our first time to be there. As a result, the two Swiss artists who had been to Zurich before identified a region of the city for us to go and explore. The area known as Limmatplatz was selected because of its scale and apparent boundaries, also, it is an area that had been through a relatively recent period of transformation and renewal. As a group we thought that this place of 'ordinary Zurich' framed with potentially something unusual taking place in the streets, would provide an interesting context for our project. Once we had a site, we then had to identify an approach? There was no doubt that it would embrace mobile cartography of the

Table 1 Schema of group members

Person	Mapping interest	Profession	Place of residence
Matthew Bissen	Walking, derive, social media integration	Architect	America
Christopher Cramer	Typically sound, ephemeral elements	Sound artist/designer	Switzerland
Claudia Schmid	Video and moving images	Film maker	Switzerland
Laurene Vaughan	Walking, theory of place, poetics	Artist/designer	Australia
Christopher Wood	Walking, sound, digital photography	Artist/graphic designer	England

Fig. 1 Limmatplatz, Zürich (Vaughan 2012)

city, and walking would be the means for exploring the locale. This was the broad
premise, and we had yet to find a means to integrate the following methods for
documenting and being in space by the group members and our individual
ambitions for the project and the workshop.

3 Insider/Outsider: Familiarity with Place

As we boarded the bus that led to the tram that would take us to the Limmatplatz
region (Fig. 1) there was a sense of being either a local or not, amongst the group.
This was an informal structure that had emerged based on nationality regarding
who would take command of navigation to the region in the city. There was an
underlying assumption that of course the two Swiss artists would know the way,
resulting in the American, the Brit and the Australian following along, in a state of
trust under the guidance of our leaders. As we journeyed we talked, sharing our
respective interests in cartography, mapping and our projects. After a short time it
became apparent that we had missed our stop and were potentially lost. It was at
this time that we realized that there were no local inhabitants in the group—we
were all outsiders who had varying levels of local knowledge. The two Swiss were
also outsiders to Zurich, and yet on the grounds of nationality, we had assumed a

familiarity with place based on proximity. This misconception seemed obvious, founded on nationality and language the international non-locals had deferred knowledge of place and map reading to those that were assumed to have local knowledge and assumed internal maps of the proposed route. The realization that we were all 'tourists' transformed the dynamic in the group—it was no longer segregated on the grounds of being local or non, and became an exchange of creative and mapping practices. Any notions of authority grounded in prior knowledge were disbanded; and the title of the project, *The Visitors* was adopted as a mascot for our individual explorations.

In our contemporary globalized world, tourists play an ever increasing role in the representation of place. Once limited to sketches in travel diaries and snap shots to share with family and friends on return to home, place as seen, documented, uploaded and published through online platforms (e.g. flickr or tripadvisor) by the tourist has expanded exponentially. Digital technologies have transformed both the places and their official discourse in image and text (Perkins 2007). Lay people (tourists) are as likely as State agencies to be the source of information about a place. And as noted by Edensor (2000, 2001) the tourist view is a particular view and an interpretation of place, especially when tourists traverse the 'heterogeneous places of the everyday' (2001). In the case of *The Visitors* we were an ad-hoc interdisciplinary group of workshop attendees, tourists by context and not by intent. We were not on holiday, or setting out on adventure outside of the boundaries of our professional work; we were undertaking a cartographic experiment, in what happened to be an unknown and unfamiliar place. Despite this difference of intention, and possible travel classification, we adopted the practices of the tourist. To any observer of the group our outsider status would easily lead to a classification of tourist; for example being lost, speaking English, not moving with the flow of the pedestrian traffic. Our methods of constant documentation whilst clasping onto paper maps as we walked, would do likewise.

3.1 Arriving in Limmatplatz

The ambition of the group was to embrace subjectivity and individuality, as an approach to creating the content of our collaborative map with out predetermining any element beyond the area being mapped. There was no need for us to focus on geographical features, or formal streetscapes and street names, the paper maps we held in our hands did all of this. Our map would provide another layer of knowing of Limmatplatz for us and for others who may observe it.

On arriving in Limmatplatz each member of the group, declared their respective approach to this cartographic exploration. We would walk individually through the streets for a period of two hours, it didn't matter if we traversed the same sections or passed each other along the way. Such overlaps in time and place were deemed to be potential features of interest in the final outcomes. The methodology that we would use can be described as the practice of noticing (Mason 2002; Vaughan

2011). This is an approach to seeing that is driven by the need to look or seek out, but rather to see what emerges from being in a place. There is a looseness to such a method of observation that embraces the possibility of serendipity, and which may result in fragmented experiences and associated data. Noticing is a process that requires the walker, the embodied cartographer, to be in place, and to endeavor to merge with the surrounding landscape as one within a context of many, to mingle with the activities of the everyday. It is a method that is far more effective within the framework of the city than rural landscapes and towns. For it is in this urbanscape that the cartographer becomes another member of the local populace. This is a populace that is in a constant state of change with people coming and going, and the cartographer's identity as outsider and observer, becomes limited to the devices that they use to way-find or document what is taking place upon the street. Different recording devices would support different kinds of gazing (Perkins 2007). As such both the subjectivity of the individual combined with the affordances of the technology, would enable us as non-cartographers, to use ad-hoc cartographic practices in the making of our map.

4 The Visitors: A Fragmented Map of Multiple Narrations of Place

As evidenced in a range of books, publications, exhibitions and events it is broadly understood across different bodies of literature from the humanities, sciences and arts that there is a growing interest in maps and mapping, of new forms of cartography, and new types of geography. In considering this, Pinder (2007) poses that '(a)wareness of the need for new ways of mapping space and places to respond to social, political and economic changes is undoubtedly fuelling cartographic imaginations' (p. 454), which in conjunction with the affordances of new technologies for map making, is the basis for this phenomenon. The outcome of this phenomenon is a new 'social life' of maps. This social life exists in both formal and informal contexts, from the inclusion of maps and mapping into products and services, to the use of maps and mapping technologies as a means for making sense of, or commenting on, the state of the world (Abrams and Hall 2006). Pinder (2007) argues that the ability to move beyond geospatial representation enables lay cartographers, who may also be known as artists, designers or tourists, to explore the world, to represent its formal and informal cultural geographies. In this way we use maps to make sense of things drawing on our visual literacy to translate a lived or observed phenomenon into a 'readable' entity, so that the map reader (who may also be named as a user or viewer) can have access to something that may be labeled as 'unseeable' due to issues of scale, distance or abstraction. It is in this way, that the map and the mapping action become sense-making activities. It should also be noted that within the arts and design there is also a counter practice of place representation whereby the aim of the mapping practice is not to make

Fig. 2 Laying out the map (Vaughan 2012)

sense, but rather to make strange. In this way the map works to expose that which is hidden or invisible with the landscape of the everyday. The mapping exploration of *The Visitors*, embraced both making sense and of making strange as approaches to documenting place.

The digital image was the dominant method used by the group to explore and to document their encounters with Limmatplatz (Fig. 2). Although there was some recording of sounds, it was captured in images as a means to becoming familiar with an unknown place. As noted by Crang (1997) it is the photographic image, which dominates the tourist view of a place—whether it is through the images taken as records of travel by individuals that are prompts for future memories, or the official representation of a place through postcards as produced by an government agency or the like. No matter whether it is private or official, in both contexts images are the records and the enablers of the experience of place. Crang claims that this is an example of the 'practices of seeing' (1997). This is a practice where the image enables people to see what is in front of them, and record what is often a predetermined view of a place. This is especially the case when such practices are coupled with a guide or guidebook. In its most extreme form of control, actual physical locations are identified and marked within the landscape to encourage tourists to record their personal experience from a predetermined site within a place. This common tourist place may be either Edensor's (2000) 'Enclavic Space' (the manufactured tourist place) and 'Heterogenous Space' (the tourist site in the midst of the everyday' or an integration of the two. For example, a viewing platform in a national park or on the side of the road, or even common locations within a street—such as Time Square in New York City. This predetermined view ensures that there is a shared dialogue between all travelers to this place, and the official record or interpretation of that place and how it wants to be perceived, can be maintained.

Crang's argument regarding the intersections between such experiences of encountering place through the touristic 'practice of seeing' highlights what would

be the normal methodology for an individual or group when walking through the streets of a city as a tourist. This however is not the most appropriate way to frame the encounter that led to the film *The Visitors*. In this project, the group deliberately set out, not to see what could be deemed to be an official view of Limmatplatz; rather, we set out with an ambition to see what we could see. The methodology guiding this enabled an approach that was less about looking and more about noticing (Vaughan 2011). Using noticing as a methodology for encountering place requires the cartographer to actively *not* seek out the view in front of them, but rather to engage with the chance sightings of the peripheral view. It is in this way, that this practice has more in common with the notion of a glimpse than a vista. To glimpse something or to have a glimpse is to have a partially distorted or obscured view. In this way the thing that is viewed by chance, becomes the prompt for either further viewing, or a partial understanding, where the rest can only be understood through either further encounter, a shift in viewing position, or by making fictitious completions or elaborations to the supposed form or meaning of that which is seen based either on memory or other data sources. To notice, to glimpse or to see out of the corner of one's eye, aspects of a place enables a different and a very subjective, understanding or interpretation of a place that is not bound or perhaps even framed by the certainties of a pre-determined view, nor the situated seeing that Crang (1997) states is the basis for much of the tourist representations of place.

In *The Visitors* project the group had identified that there was a shared interest in subjective mapping and informal cartographies of the everyday, but the form of our individual mapping outcomes from our respective creative practices was less defined. We had an identified site (Limmatplatz), a methodology (walking and noticing), a structure (2 hour and any form of documentation that you prefer or seems appropriate) and a political premise (equality in that no one persons' mapping method would dominate the form or aesthetic of the outcome). Our method would be to accrue data, which we would then analyse, synthesise and collate into a collaborative cartographic representation of one place. Conversation and cultural exchange amongst the various participants underpinned the collaboration, and these were the unnamed and yet densely present methods that would ensure and support the cartographic outcome. It could be argued that we were non-tourists masquerading as flaneurs, randomly yet deliberately wandering, with a clear intention of gathering data that would ultimately be integrated to create a subjective map.

After the two hours of encountering Limmatplatz had passed the group reconvened in a cafe at the point where we had all departed. It was here that there were excited exchanges about what had been seen or heard. What was noticed was noted, images were shared on small screens, sound bites were listened to, and fingers outlined paths on the official tourist maps. It was apparent that on the whole we had traversed five separate paths within the limited distance of Limmatplatz, and that there had moments of sighting of each other and overlaps in some of our wanderings. As we scoured through the documentation two interesting things began to emerge. Firstly there were five distinct views and aesthetics of

representation, and secondly, there were patterns of attraction in terms of colours, architectural features or anomalies on the streetscape. As we examined the documentation new themes of intersection began to emerge, leading ponderings of what something may mean. The boundaries between phenomenon, location, representation and meaning were in a constant state of flux and evaluation.

To return to Crang (1997) and his claim that '[t]he picture is not simply a translation of an ideology on to celluloid, nor is it simply an expression of a momentary experience… [photography] is not just a picturing of landscape, nor representation of places—it is seizing a moment in a place. It is communicating some point about experience in one particular place and time to an audience or viewer in another place and time' (p. 367). In *The Visitors* project, the group had decided to use photographic imagery as a means to document a place, and to explore the relationships and possibilities for a lay, or non-cartographic yet creative realisation of a map. It would be a composite map of five individual views integrated into one form (Fig. 3). As an interdisciplinary manifestation of what Corner (1999) would describe as a 'layering' of place. Although Corner is referencing a process for the realization of a new site *in* place, the intention of layering separate views or components one upon another as a means to represent the complexity that *is* a place, is consistent with the experiential representation map of a place that we created. *The Visitors* was a map of the informal geographies of phenomenon that are noticed, but not sought out. In this way, the map that was realised through this emergent methodology was a map that seized a moment in a place, and that could be viewed by others in another time and place. It embraced Crang's definition of the affordances of the photographic image, and it did this with the ambition to be seen as a cartographic outcome that would narrate experience and qualities of a place.

Fig. 3 The Visitors an imprecise map (Vaughan 2012)

4.1 Making the Map

Grounded in this approach of mapping through layering the group decided to exploit the affordances of the photographic image, its qualities and unique view and to create a topological landscape of the region through a scattering of all the images taken by the team of non-cartographers. These were then laid out upon a surface of a table and like all maps decisions about form, priority and intention had to be made. Aesthetically it was decided that when printed, the images would retain the white border that harks back to celluloid images rather than digital prints. This would be a border that both marked the boundaries of one view, or one moment as seen by the photographer; and the border as a consistent aesthetic device that would unite the 100s of images as they lay on the plain of the table (Fig. 3). This was phase one of the integrated map of *The Visitors*.

Phase two involved narrating the trajectories of individual passages through Limmatplatz. Each group member identified within the many images that were spread on a plain some of their own images, they then laid them out on top and turned them over to make them obvious within the messiness of the whole. They then integrated their individual tourist maps featuring marked lines of their walking into the tableau. Then under the watchful eye of the vidoeographer, we each then drew the lines of our walking, the images of our noticing one after another as they traced the trajectory using the viewfinder as a second order walk in the place; with this the composite map became the shared narrative of the five whilst also noting the uniqueness of the one. There were five distinct views united as one shared and collective interpretation and representation of Limmatplatz in Zurich, on June 12, 2012. The method that was used shares elements with the practices of ethnographic filmmaking (Pink 2007), where the ethnographer walks with the subject as they co-record through action and narrative the nature and experience of place. However in this case there are two walks and multiple narratives. There were the initial walks by the participants, documenting their own noticings of an unfamiliar place. Then there was a secondary walk, realized through a finger tracing across a landscape of images marking out the line of an embodied walk. In this way multiple orders documentation are represented, and the first and second order narratives are conveyed. Through this process the unknown was transformed to being slightly familiar, in the same way that a tourist recounts a narrative of an encountered place through anecdotal observations and conversations. In this case one of the many heterogeneous spaces in the city of Zurich, was transformed to be an imprecise geography of an accidental touristic encounter.

5 Conclusion

The Visitors was a cartographic exploration of a collaborative approach to shared and individual narratives of place. It embraced the nuances of the tourist whilst also rejecting the fundamental premise of the controlled or formal tourist view.

The aim was to explore or create an official interpretation of Limmatplatz, but rather to embrace subversive, individual meanderings in place whilst being directed by the formalities of a tourist map. Through poetic devices of abstraction and aesthetics *The Visitors* challenged notions of authority and the collation of data as a means to synthesise and create a whole. In this project we celebrated diversity, we were challenged about notions of leadership and familiarity grounded in the distances of nationality or professional practice. This project was a touristic cartographic exploration of Limmatplatz and the 100s of images that it manifest created the embodied map that *The Visitors*—a filmic translation five individual members, into one fragmented narrative of place.

References

Abrams J, Hall P (eds) (2006) Elsewhere mapping. University of Minnesota Press, Minneapolis

Bissen M (2012) A personal geography. Cartography and narratives. http://cartonarratives.wordpress.com/projects. Accessed 3 Nov 2012

Cacquard S (2011) Cartography 1: mapping narrative cartography. Progress in human geography, Sage Publications, pp 1–10. http://phg.sagpub.com. Accessed on 7 Nov 2011

Casey E (2001) Between geography and philosophy: what does it mean to be in the place-word? Ann Assoc Am Geogr 91(4)2001:683–693

Corner J (1999) The agency of mapping: speculation, critique and invention. In: Cosgrove D (ed) Mappings. Rearton Books, London, pp 213–242

Crang M (1997) Picturing practices: research through the tourist gaze. Prog Hum Geogr 21(3):359–373

Edensor T (2000) Staging tourism, tourists as performers. Ann Tour Res 27(2):322–344

Edensor T (2001) Performing tourism, staging tourism. Tour Stud 1(1):59–81

Gill D (2012) Erratic space. Cartography and narratives. http://cartonarratives.wordpress.com/projects/. Accessed 3 Nov 2012

Lippard L (1997) Lure of the local. The New Press, New York

Mason J (2002) Researching your own practice. The discipline of noticing. Routledge, London

Perkins C (2007) Photography. In: Douglas I, Huggett R, Perkins C (eds) Companion encyclopaedia of geography. Routledge, London, pp 587–602

Piatti B, Bär HR, Reuschel A-K, Hurni L, Cartwright W (2009) Mapping literature: towards a geography of fiction. In: Cartwright W, Gartner G, Lehn A (eds) Art and cartography. Springer, Berlin, pp 1–16

Pinder D (2007) Cartographies unbound. Cult Geogr 14(3):453–462

Pink S (2007) Walking with video. Vis Stud 22(3):240–252

Vaughan L (2011) Roaming montreal: seeking the representation of the 'geographic self. In: Caquard S, Vaughan L, Cartwright W (eds) Mapping environmental issues in the city: cartography and arts cross perspectives. Springer, Berlin, pp 146–159

Vaughan L (2012) Telling place: narratives of an unknown city. Cartography and narratives. http://cartonarratives.wordpress.com/projects/. Accessed 3 Nov 2012

Watson C (2012) Hyper real and narrative maps. Cartography and narratives. http://cartonarratives.wordpress.com/projects/. Accessed 3 Nov 2012

Wood J (2012) Personal cartography. Cartography and narratives. http://cartonarratives.wordpress.com/projects/. Accessed 3 Nov 2012

Towards a Spatial Analysis of Toponym Endings

Tobias Dahinden

Abstract The target of this article is to define and use a statistical measure to determine endings of place names. The definition of 'ending' is based on the occurrence of a certain end-string in a gazetteer. Based on this definition a part of the GeoNames-gazetteer is analysed in respect to detect and rank possible endings. The spatial distributions of the most outstanding endings are presented.

Keywords Toponymy · Analysis · Detection · Gazetteer · Vernacular

1 Introduction

Toponyms consist of a single name (simplex), composition and move together (compound), and derivation (derivata). Similarities between toponyms are obvious: some places use the same name (e.g. *Cambridge*) and, at least in Germanic languages, some names of places have the same ending.

Concerning endings, it is possible to differ between suffixes (e.g. *-nia* in *California* and *Pennsylvania*) and primary words (e.g. *-land* in *Greenland* or *Maryland*).

The aim of this article is to detect salient endings of Toponyms in a statistical way, and to present the source area for some of them. The article has been organized in the following way: in the Sect. 2 some knowledge about endings is presented. A statistical definition of 'ending' is given in Sect. 3. The Sect. 4 gives a brief overview on the gazetteer that is used in this research. Some results are presented in Sect. 5, i.e. endings that are salient according to the definition in

T. Dahinden (✉)
Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover, Hanover, Germany
e-mail: tobias.dahinden@ikg.uni-hannover.de

Sect. 3 and the source area of some endings are shown. The Sect. 6 gives an outlook on future work.

2 Expectations about Endings

A considerable amount of literature has been published about names and endings of names, mainly in linguistics. In general, an author investigates a small area and analyses the names based on the meaning of parts of the names. As an example, Andrießen (1991) traced for Hessen (Federal state in Germany) *"-berg, -hagen, -hausen, -inghausen, -rode* [and] *-sen (< -hausen)"* as common endings for place names. Waser (2002) reported on place names from the German-speaking part of Switzerland. He discussed the diminutives *-i, -etli, -ti, -li, -ili, -eli, -dli* in place names. According to a publication of Bickel (1998) the common endings of toponyms for north-west Switzerland are *-walen, -ingen, -berg, -wil, -inghofen, -heim,* and *-dorf*.

From the referred research it can be noted that

- the size of the endings differs from 1 to 9 characters (with a mean value of 4),
- endings are typical for certain regions,
- some endings are a superset of another ending, such as *-sen* for *-hausen*.

A long but unproved list with endings consisting on suffixes and primary words is published on Wikipedia (2012).

Names may have a misleading meaning in the current spoken language. According to Bossong (2002) toponyms are usually adapted to the current language although they are often much older. He mentioned a place called *Bischofsheim*[1] in the Vosges Mountains were at no time a "Bischof" was present. He mentioned that there are similarities to the toponym *Bizcoy* (a place near Alicante in Spain). A related toponym is *Biscofesheim*, which is the old name for *Tauberbischofsheim* (Wikipedia 2009). So, the hyphenation of *Bischofsheim* should be rather *Bisc-hof-(s)-heim* than *Bi-schof-(s)-heim*. It would be too easy to compare the names with a dictionary of the language of this area.

The purpose of this research is to find patterns in endings of toponyms in a gazetteer with entries that are widely distributed. Because analyses on endings are usually for a rather small area it is not satisfying just to apply the known endings of some areas to a larger region. It is also not sufficient to use a dictionary to find primary words because names are sometimes much older than the spoken language.

Thus, we suggest using a statistical method together with a gazetteer instead of using an existing list of common endings. Unfortunately, there is no statistical definition for endings according to our knowledge.

---

[1] Translated: the Bishop's home.

3 A Statistical Definition for Endings

As a simplification for this article, the term ending is used for a suffix as well as for a primary word or a compound. To determine common endings in a gazetteer a statistical definition is unavoidable.

An ending could be defined as something like an end-string s of a word, which occurs more often than other end-strings. It has to be taken into account that longer strings are less often than short strings because a long string is always a subset of a short string. Hence, endings cannot be defined by the number of occurrence only; otherwise an end-string like *-orf* would be a more prominent ending than the primary word *-dorf*.

The definition might be based on the idea of checking weather adding an additional character a certain substring could be random or systematic. Thus, the occurrence of a character c in front of a string s is analysed. This is done using the ratio R of the number of words with the ending $c + s$ and the number of words with ending s. Shortly written: $\#(c + s)/\#s = R$.

The chance C of an end-string to be an ending is regarded as the sum of the occurrence-ratio R of the letters divided by the length of the substring.

Def. An ending is an end-string s where the chance C of being an ending is greater than C of all substrings, which are longer or shorter by one character.

As an example the string *-nplanggen* is discussed. It's the most often occurring end-string in a gazetteer called *Urner Namenbuch*[2] (Hug and Weibel 1988–1991). The end-string *-nplanggen* consists of nine characters and thus, can be subdivided in nine elements (c.f. Table 1). It occurs in 130 names while *-planggen* occurs in 534 names. This leads to $R = 0.24$. This means that 24 % of the words with the substring *-planggen* also have the substring -nplanggen. The substring *-langgen* occurs as often as *-planngen*, thus R is 1.

The chance C for *-nplanngen* to be an ending is the sum of all ratios R ($5.97 = 0.24 + 1 + \ldots + 0.26$) divided by the number of characters (9). An analysis of Table 1 shows that a maximum for C occurs for the string *-planggen* and for *-en*. Actually, the term *planggen* is a primary word (derived from latin *planca*) in the language represented in the gazetteer.

4 Gazetteer and Research Restrictions

This study makes intensive use of the GeoNames gazetteer (Wick 2012). The gazetteer is based on more than 60 different data sources such as Wikipedia's Wikipedia-World, U.S. Geological Survey's Geographic Names Information System, and Swisstopo's Swissnames. It contains more than 8.2 million

---

[2] Digital access on: www.ortsnamen.ch.

Table 1 Division of -*nplanngen* in possible endings: R gives the ratio of counts of a word with its predecessor. C is the sum of R of the letter and all predecessors divided by the length of the string, large values indicate that it is very improbable that this term occurred by accident

Ending	Counts	R	Sum(R)	Length	C
-nplanggen	130	0.24	5.97	9	0.66
-*planggen*	534	1	5.73	8	0.72
-langgen	534	1	4.73	7	0.68
-anggen	534	0.95	3.73	6	0.62
-nggen	562	0.73	2.78	5	0.56
-ggen	762	0.71	2.04	4	0.51
-gen	1067	0.16	1.32	3	0.44
-*en*	6596	0.91	1.17	2	0.58
-n	7288	0.26	0.26	1	0.26

geographical names; it is published using the Creative-Commons Attribution 3.0 License. It stores for each place a name, a latitude and longitude in WGS84, an id, a name in ASCII, a list with alternative names, information about the feature type, population, the code about the administrative division, and the elevation.

Some names in the gazetteer consist of more than one word (such as "Moreira de Cónegos" or "Museum of Modern Art"). In such cases it is not possible to know a priori which part should be analysed. Thus, for the study the names were restricted to consist of single words.

The investigation area is restricted to Germany, Austria, and Switzerland. The majority of names are from an area where German is spoken. This means that the research is restricted to agglutinative languages; these are languages where compounds are set together. Thus, the words can be very long.

These restrictions shortened the list with places to 178032 names. The mean density of the entries in the database is around one name per 2 km^2 for this area.

The analysis was performed on the names written in ASCII, no phonetics was used. Thus, Umlauts were not taken into account,[3] two very common digraphs (*sc,*[4] *ch*) were handled as one single character; other digraphs (*ck, ie, ph, tz,*…) are handled as two characters. The length of the endings is restricted to 6 letters to reduce the runtime of the analysis.

The result is restricted to substrings that occur at least 100 times in the gazetteer. Still, there are 1013 different end-strings to be analyzed.

---

[3] Unfortunately, Umlauts such as *ü* were stored inconsistent, partly as *u* and partly *ue*.

[4] This is related to *sch*.

Fig. 1 Distribution of places with a name ending on -*berg*
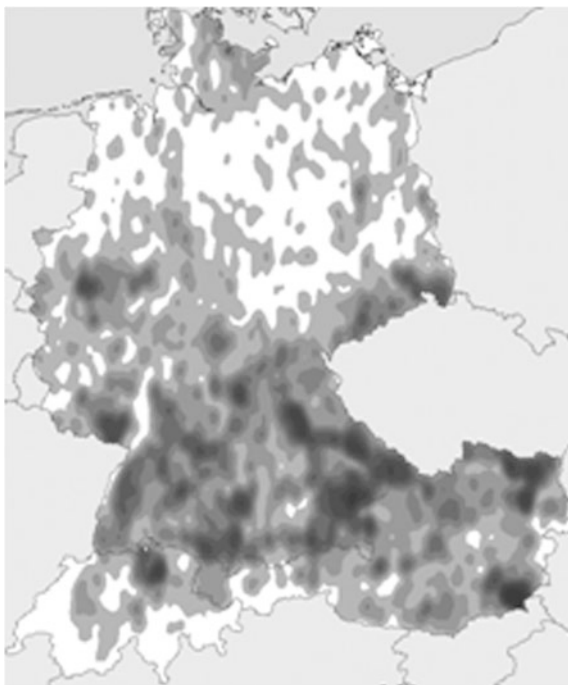
5 Analysis

Using the definition of Sect. 3, the end-strings of the place names are classified and ranked. The 12 strings with the highest chance C to be an ending are: -*berg, -spitze, -enberg, -hausen, -dorf, -weiler, -endorf, -erberg, -kreuz, -graben, -enkamp, -heim*.[5] Comparing this list with literature about endings (c.f. Sect. 2), it becomes obvious that the evaluation method privileges primary words and compounds of -*en* or -*er* with a primary word, e.g. -*kamp*.

Figures 1, 2, 3, 4 show the probability density function of places with the ending -*berg*, -*dorf*, -*weiler*, and -*enkamp*. The images are based on kernel density estimation. All of them have been calculated using an Epanechnikow kernel (ESRI Support Centre 2010) with a bandwidth of 0.2°. The bandwidth has been established by a visual observation.

As one can see from the Figures the distribution varies for different endings. While places ending on -*berg* exist nearly everywhere in the investigation area, places ending on -*dorf* are common in the eastern part, -*weiler* is common in the shout-west and -*enkamp* seems to be typical for the north-west of Germany.

---

[5] A longer list is presented in the appendix of this article.

Fig. 2 Place names ending on *-dorf* occur more frequently in the eastern part of the investigation area
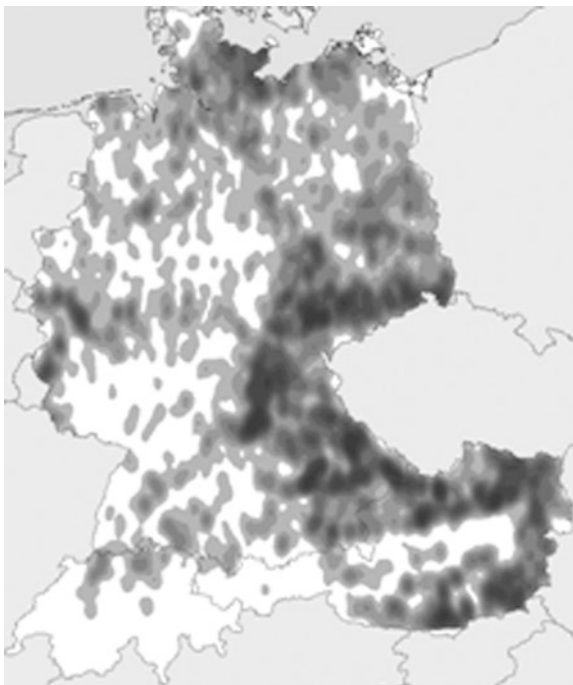
Fig. 3 The ending *-weiler* seems to be related to the south-west

Fig. 4 The end-string -*enkamp*, that is a compound of the suffix -*en* and the primary word -*kamp*, is common in north-west Germany

Figures 5 and 6 show the density estimation of 42 endings. The estimation is based on a Nadaraja-Watson type kernel smoother (Baddeley and Turner 2005, Diggle and Arnold 2003) with a bandwidth of 1°. The endings were selected according to *C*. Compounds like -*enkamp* are not taken into account for the presentation. In Fig. 5 the colour scale is the same for all images while it changes in Fig. 6.

Figure 5 is useful to uncover endings that dominate a region, noticable are: -*au*, -*bach*, -*berg*, -*dorf*, -*feld*, -*hausen*, -*hof*, -*ingen,* and -*ow*. Unsurprisingly, these are the same endings that occur most often in the database.

Figure 6 is useful to see, where the endings are probably located. Comparing the single images correlation between endings could be guessed, e.g. there might be a correlation between -*horst*, -*moor*, and -*stedt*. What is surprising is that there could be a kind of negative correlation of -*dorf, -hausen,* and (maybe) -*heim*.

Figure 7 provides the distribution of -*ingen.* The image was calculated using adaptive kernel density estimation (Silverman 1986, Baddeley and Turner 2005) with 20 repetitions. From the Figure it becomes obvious that this ending is present mainly in the South but also in northern regions.

Comparing the endings to dialect maps lead to some correlation between the endings and the dialects, e.g. -*ingen* is often used where people speak Alemannic German (south west) but it is also appearing in northern parts of Germany. On the
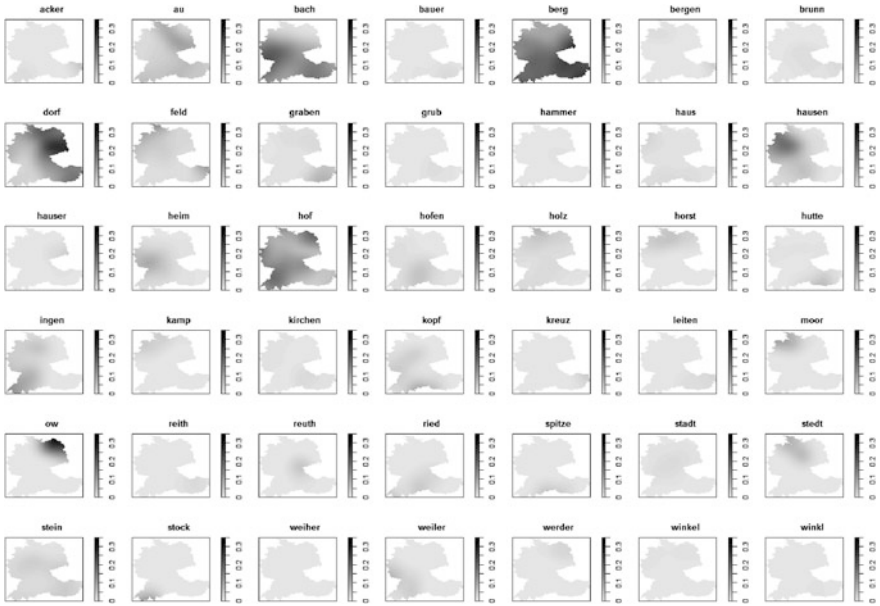
Fig. 5 Density of 42 different endings (the darker the area, the more probable an ending occurs). All images are displayed using the same scale
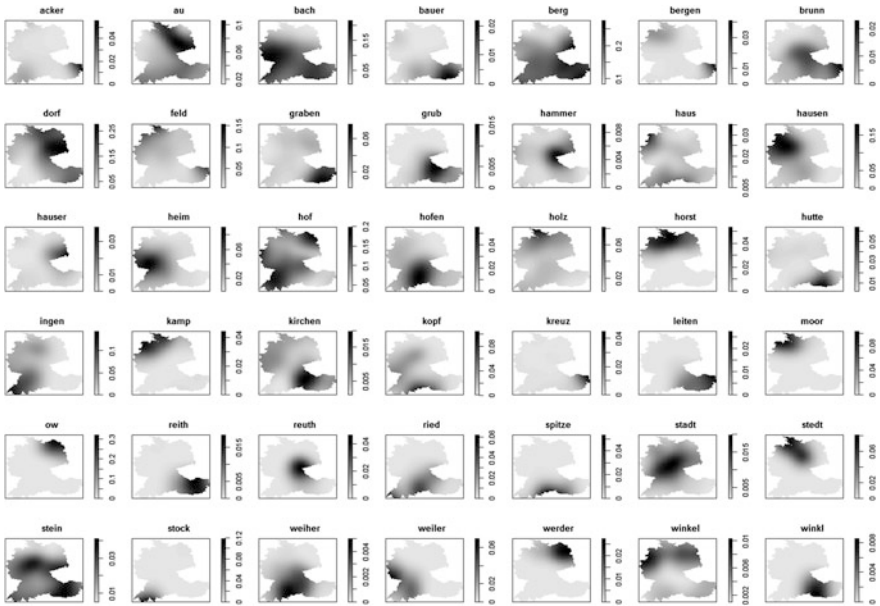
Fig. 6 Density of 42 different endings (the darker the area, the more probable an ending occurs). Note: The images use different scales
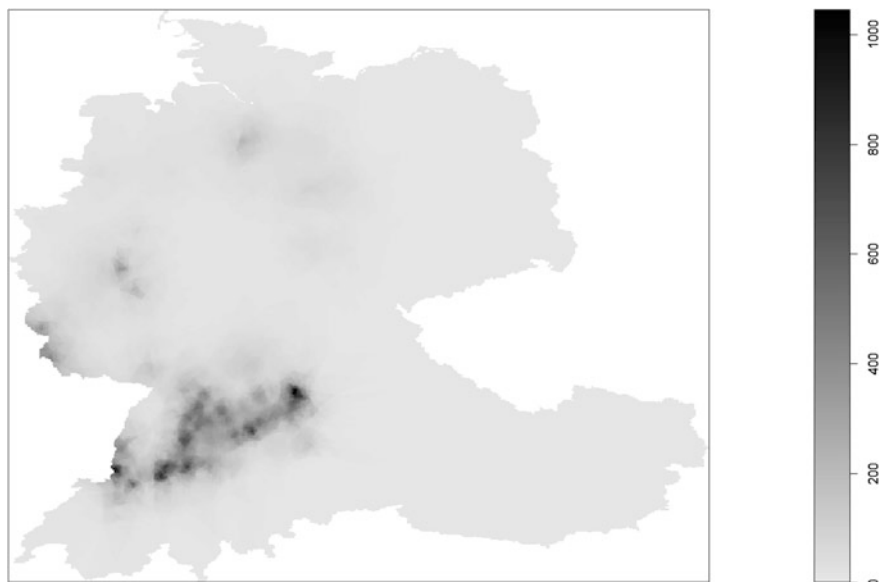
Fig. 7 Probability to find places ending on *-ingen* (the darker the area, the more probable the ending occurs). The image has been calculated using adaptive kernel density estimation

other hand, there might be a correlation between *-horst, -moor, -stedt* and Lower German.

6 Conclusion and Further Work

In the article the term 'ending' has been defined using a measure based on the absolute number of end-strings in a database and the length of the string. The measure is rather simple; e.g. it does not take into account that letters have a certain frequency.

The definition of 'ending' does not distinguish between primary words, suffixes, simplex, and compound. Future studies are recommended to deal with the splitting of endings. This would involve the use of a dictionary with primary words. The dictionary itself could be established using the definition of this paper. On the same time, endings should not be restricted in length, unlike it has been done here. Further, the analysis should be enlarged to languages where words tend to be isolated. This implies that the analysis must not be restricted to endings but it has to deal with any kind of patterns in words.

For the analysis of the place names the investigation area was restricted to German-speaking countries and only ASCII names of the GeoNames-database were used. A more profound study could deal with phonetic algorithms such as

Soundex (c.f. Knuth 1973) or *Kölner Phonetik* (Postel 1969). Using such a method would allow bundling similar words. The German *-heim* and the English *-ham* or the German *-burg* and the French *-bourg* would be handled correctly as the same word. Similar the German *-reuth, -rode, -reit, -rütti, -ried, -roda,...* would be seen as one single expression. But undesired friends may also arise: the Kölner Phonetik for *Othenbruch, Osnabrück, Oschenberg* is *O86174*, meaning the words are the same according to the algorithm, but most probably they are not. A solution would be the use of the International Phonetic Alphabet as a base for an analysis.

The analysis was performed on a small part of the GeoNames-database and a ranking of possible endings of place names has been created based on the definition in Sect. 3. The spatial distribution of the most-outstanding endings has been shown. A visual analysis of the distribution suggested weak correlation between endings, possibly also negative correlation. The current study was unable to make a profound analysis on spatial point patterns. It is suggested that a detailed statistical analysis on the distribution is applied. A major issue would be the un-equal distribution of the places in space and to run a multivariate point pattern analysis related to the end-strings.

Further studies on the current topic should deal with larger investigation regions and/or more detailed gazetteers. Research questions that could be asked include the existence of ending structures in other languages, the correlation of patterns in place names and the spatial distribution of these names, the analysis of compounds, the search for pattern in parts of the names, and an analysis of names consisting of more than one word.

A.1 7   Appendix

The 150 salient endings evaluated for this article (beside those starting with *-en, -er*, or *–es*):

-berg, -spitze, -hausen, -dorf, -weiler, -kreuz, -graben, -heim, -weiher, -kamp, -hauser, -bach, -grub, -elberg, -stein, -reuth, -schlag, -winkl, -acker, -holz, -chberg, -scheid, -hammer, -feld, -neuhof, -werder, -brunn, -nsdorf, -hofen, -stedt, -kopf, -reith, -stadt, -ndberg, -winkel, -tzberg, -haus, -horst, -rnberg, -ingen, -hof, -bauer, -stock, -onberg, -leiten, -ried, -au, -msdorf, -nsberg, -bergen, -gsdorf, -ow, -hutte, -eldorf, -moor, -ssberg, -strass, -garten, -isberg, -undorf, -storf, -indorf, -tsberg, -trup, -itz, -dsberg, -amberg, -wiesen, -hlberg, -spitz, -krug, -heide, -bruck, -moos, -holzen, -lsheim, -brink, -llberg, -steig, -itze, -burg, -schach, -grund, -berge, -muhle, -aubach, -koog, -kogel, -elbach, -schutz, -usen, -wald, -chholz, -bronn, -sgrun, -en, -kanal, -hagen, -hardt, -grat, -statt, -chfeld, -reute, -heid, -tzbach, -hofe, -reit, -iler, -hlag, -thal, -er, -buhel, -muhlen, -lehen, -wiese, -felden, -halden, -inbach, -bichl, -ppach, -bruch, -nfels, -rnbach, -riegel, -eck, -roth, -mmer, -sleben, -imbach, -elhof, -sattel, -ikon, -etten, -aben, -ambach, -buttel, -hafen, -hutten, -ewitz, -oven, -ing, -harte, -ssbach, -hub, -asser, -hlbach, -roda, -acher, -furth, -stall

References

Andrießen K (1991) Siedlungsnamen in Hessen. Verbreitung und Entfaltung bis 1200. Deutsche Dialektgeographie. N.G. Elwert, Marburg

Baddeley A, Turner R (2005) Spatstat: An R package for analyzing spatial point patterns. J Stat Softw 12(6):1–42

Bickel H (1998) Ortsnamen als Quellen für die Siedlungsgeschichte am Beispiel der Nordwestschweiz. XIXth international congress of onomastic sciences, Aberdeen, 4–11 Aug 1996

Bossong G (2002) Der Name Al-Andalus: neue Überlegungen zu einem alten Problem. In: Restle D, Zaefferer D (eds) Sounds and systems. Studies in Structure and Change. A Festschrift for Theo Vennemann. Mouton de Gruyter, Berlin, pp 149–164

Diggle PJ, Arnold H (2003) Statistical analysis of spatial point patterns. Arnold, London

ESRI Support Centre (2010) ArcGIS desktop help: kernel density. Environmental Systems Research Institute, Inc. http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=6190&pid=6188&topicname=Kernel_Density. Accessed 13 April 2010

Hug A, Weibel V (1988-1991) Urner Namenbuch: Die Orts- und Flurnamen des Kantons Uri. 4 vols. Altdorf

Knuth DE (1973) The art of computer programming: volume 3. Sorting and searching. Addison-Wesley, Reading

Postel HJ (1969) Die Kölner Phonetik - Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse. IBM-Nachrichten 19:925–931

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall (CRC Press), London

Waser E (2002) Das Diminutiv in Orts- und Flurnamen. Congreso Internacional de Ciencias Onomásticas, Santiago de Compostela

Wick M (2012) GeoNames. www.geonames.org. Accessed 8 Nov 2012

Wikipedia (2009) Tauberbischofsheim. Wikipedia Die freie Enzyklopäsie, Mediawiki Foundation. http://de.wikipedia.org/wiki/Tauberbischofsheim. Accessed 4 Oct 2009

Wikipedia (2012) German placename etymology. Wikipedia the free encyclopedia, Mediawiki Foundation. http://en.wikipedia.org/wiki/German_placename_etymology. Accessed 14 Nov 2012

A Contextual ICA Stakeholder Model Approach for the Namibian Spatial Data Infrastructure (NamSDI)

Kisco M. Sinvula, Serena Coetzee, Antony K. Cooper, Emma Nangolo, Wiafe Owusu-Banahene, Victoria Rautenbach and Martin Hipondoka

Abstract In 2011, the Namibian parliament presented and promulgated the Namibian Spatial Data Infrastructure (NamSDI) with the aim of promoting the sharing and improved access and use of geospatial data and services across Namibia. Notable SDI models, developed from the enterprise, information and computational viewpoints of the Reference Model for Open Distributed Processing (RM-ODP), comprise direct and indirect roles of stakeholders and special cases of each general role in an SDI. Hence, the International Cartographic Association (ICA) model was used to identify the stakeholders in and around NamSDI, which is still at the infancy stage of development. The application of a high-level ICA model proved to be relevant and useful in discriminating and categorizing NamSDI stakeholders according to their roles and vested interests. Some stakeholders, such as official government mapping agencies, assume multiple roles, while others, such as database administrators, are not yet active. In the absence of baseline data and given the infancy status of NamSDI, attributes such as skills, capacity of producers and service providers, were not considered. Modelling NamSDI stakeholders in the context of ICA's stakeholder model contributed significantly to a better understanding of NamSDI stakeholder types and subtypes and pointed out gaps that may hinder its successful and effective implementation.

K. M. Sinvula (✉) · S. Coetzee · A. K. Cooper · W. Owusu-Banahene · V. Rautenbach
Centre for Geoinformation Science, University of Pretoria, Pretoria 0002, South Africa
e-mail: ksinvula@gmail.com

A. K. Cooper
Built Environment, CSIR PO Box 395Pretoria 0001, South Africa

E. Nangolo
Independent Researcher, Windhoek, Namibia

M. Hipondoka
Department of Geography, History and Environmental Studies, University of Namibia, Windhoek, Namibia

1 Introduction

A *spatial data infrastructure (SDI)* is an evolving concept for facilitating, coor-
dinating and monitoring the exchange and sharing of geospatial data and services,
and the metadata about both. It encompasses stakeholders from different levels of
representation (local, regional and national) and disciplines. An SDI is more than
just the technology of a *geographical information system (GIS)*: it is a collection of
technologies, policies and institutional arrangements and provides the basis for the
discovery, evaluation and application of geospatial data and services (Cooper et al.
2013a, b), adapted from Hjelmager et al. (2008) and Nebert (2004).

Namibia, with a land surface of 825,418 km^2, is the 34th largest country in the
world, and has a population of about 2.1 million (Fig. 1).

During 2011, the Statistics Act (No 9 of 2011) (NPC 2011a), the Draft National
Spatial Data Infrastructure Policy of Namibia (NPC 2011b) and the National
Spatial Data Infrastructure Standards Schedule of Namibia (NSA 2011) brought
into being NamSDI. The Commission on Geoinformation Infrastructures and
Standards of the International Cartographic Association (ICA) has been using the
RM-ODP (ISO/IEC 10746-1:1998) to develop formal models of an SDI from the
enterprise and *information viewpoints* of RM ODP (Hjelmager et al. 2008), and
from the *computational viewpoint* (Cooper et al. 2013a, b). These viewpoints
contribute towards a more holistic interpretation of an SDI, independent of specific
SDI legislation, technology and implementations (Cooper et al. 2013a, b). A key
part of these models is the identification of general roles of stakeholders within and
around an SDI: *Policy Maker*, *Producer*, *Provider*, *Broker*, *Value-added Reseller
(VAR)* and *End User* (Hjelmager et al. 2008). The Commission also identified 37
special cases of these general roles (Cooper et al. 2011).

As NamSDI is still in its early days of development, it was useful to apply the
Commission's models and identify the stakeholders within and around NamSDI,
their general roles and the special cases of these general roles, as a contribution to
the development of NamSDI. The community of stakeholders for NamSDI is
relatively small, concentrated and has a high degree of acquaintances among them.
Sinvula et al. (2012) explored and established that the composition of NamSDI
corresponded with the enterprise viewpoint of the RM-ODP in terms of purpose,
objectives and policies as documented in NamSDI legal frameworks.

This chapter presents NamSDI stakeholders as sub-types of the stakeholder
types as identified in the ICA model. The remainder of the chapter is structured as
follows: Sect. 2 provides background information on NamSDI and the models
used; Sect. 3 explains the methodology followed to identify NamSDI stakeholders;
in Sect. 4 NamSDI stakeholders are presented; Sect. 5 discusses the results; and
Sect. 6 concludes.

Fig. 1 Map of Namibia in Africa and the world (Wikipedia 2012)

2 The Namibian Spatial Data Infrastructure

The geospatial community of Namibia recognized the common usage of fundamental datasets (base data) by different users across various applications. Sections 47 and 48 under Part 9 of the Statistics Act (No 9 of 2011), established the NamSDI. The Statistician-General of the Namibia Statistics Agency, in consultation with the CSD, issued the schedule for standards relating to NamSDI, in terms of Section 36(2) of the Statistics Act (No. 9 of 2011) (NSA 2011).

The successful implementation of SDIs is reliant on relevant expertise, which Namibia lacks. Based on the authors' observations and experience, Namibia has very few people with a clear and systematic understanding of SDIs. The gap in expertise has implications on the implementation of NamSDI.

Many stakeholders from various organizations in Namibia constituted a team that played a major role in initiating and currently implementing NamSDI. A contextual analysis of these stakeholders improves the understanding of NamSDI, which is required in shaping its future evolution.

3 Methodology

Hjelmager et al. (2008) identified and described six stakeholders in the enterprise viewpoint of an SDI using UML use case diagrams (Fig. 2), and recognized that an individual stakeholder can execute different roles. For example, an organization can act as a *policy maker*, who sets out rules and policies for an SDI, and at the same time, be a *producer* of data and services required in an SDI. Cooper et al. (2011) discussed these stakeholders from the perspective of volunteered geographical information (VGI), but also took this further by identifying various special cases of these general roles termed 'sub-types'.

The ICA's model describes the characteristics of an SDI at a high level of abstraction. The objective here is to model NamSDI stakeholders as subtypes of the ICA's model stakeholders, i.e. in practice. Such a modelling exercise improves the understanding of NamSDI, but also serves to test the behavior and applicability of the abstract model to specific SDI instances.

The ICA's stakeholder features and patterns (scientific and theoretical in nature) were associated with NamSDI stakeholders and their roles (observed or operational). This procedure included direct observation of NamSDI in the form of impressions, literature and formal objective measurement. NamSDI stakeholders were first identified at a workshop of the South Africa/Namibia bilateral project with representation from Namibia, members of the ICA Commission and other scholars who reviewed and confirmed the validity of the facts regarding the description and definition of NamSDI in terms of the ICA model.

4 NamSDI Stakeholders

Hjelmager et al. (2008) identified and defined six individual stakeholders from the ICA model (Table 1).

4.1 Policy Maker

The Government of the Republic of Namibia (GRN) is the notable *policy maker* and *producer* of fundamental (base/reference) spatial datasets, through various line ministries and state owned agencies (e.g. National Planning Commission (NPC), Ministry of Lands, Namibia Statistics Agency).

The legal and policy framework of NamSDI was approved and promulgated by the Namibian Parliament in their role as *legislators* (a sub-type stakeholder of policy marker). The legislators played a key role in determining the framework within which NamSDI had to exist (NPC 2011a). The NamSDI legislation was integrated in the Statistics Bill that was initiated and prepared by the Office of the
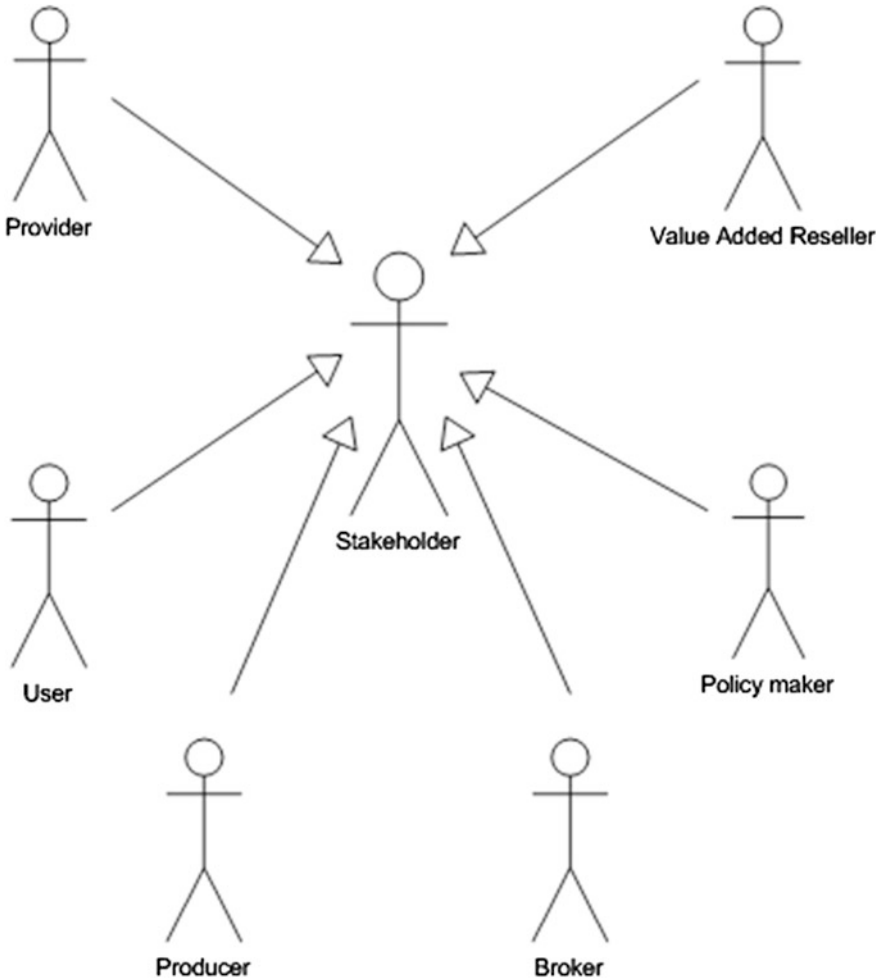
Fig. 2 The ICA's stakeholders from the enterprise viewpoint of RM-ODP (Hjelmager et al. 2008)

Table 1 Types of ICA model stakeholders (Hjelmager et al. 2008, Cooper et al. 2011)

Stakeholder	Description
Policy maker	*A stakeholder who sets the policy pursued by an SDI and all its stakeholders*
Producer	*A stakeholder who produces SDI data or services*
Provider	*A stakeholder who provides data or services to users through an SDI*
Broker	*A stakeholder who brings users and providers together and assists in the negotiation of contracts between them*
Value-added reseller (VAR)	*A stakeholder who adds some new feature to an existing product or group of products, and then makes it available as a new product*
End user	*A stakeholder who uses the SDI for its intended purpose*

President (OP) through NPC and the Ministry of Lands, Resettlement and Reha-bilitation (Table 2). The Minister for Presidential Affairs under OP initiated and tabled the bill in parliament for approval after consultation with the Ministry of Justice.

Table 2 NamSDI stakeholder types and sub-types

The *decision maker* for the NamSDI involves the Committee for Spatial Data (CSD), whose role is to advise the Minister and NSA and consists of members appointed in accordance with the Act (i.e. Statistician General and others) under the chairmanship of the Surveyor General (SG). The NamSDI was initiated by the geospatial community which consisted of members from the office of the SG, the Central Bureau of Statistics under NPC and other government ministries. The

Stakeholder			Examples
Type	Sub-type	Sub-sub-type	
Policy Maker	Legislator		▪ NPC within the Office of The President (OP) prepared the Act in consultation with line ministries ▪ Different levels of government were consulted (they are not legislators) and GIS community. ▪ Min. of Justice assisted with legal drafting (polished). ▪ OP presented the Act before Parliament. of Namibia (PoN) ▪ National Council (NC) reviewed the Act. ▪ PoN modified and approved the Act.

	Decision maker		▪ Committee for Spatial Data (CSD)
			  ✓ Advises the Minister and NSA
			  ✓ Members as per Act include Statistician General (St.G) and others
			  ✓ CSD is chaired by the Surveyor General (SG)
			▪ Ministry of Lands and Resettlement (MLR)
			  ✓ Implement own standards without consulting CSD
			▪ Minister responsible for Statistics (Minister for Presidential Affairs within OP)
			▪ NSA through St.G administers the Act

	Secretariat		NamSDI Secretariat within NSA

	Champion		▪ Individuals at various organizations
			  ✓ NPC (before the Act)
			  ✓ NSA (after the Act)
			  ✓ MLR
			  ✓ Ministry of Environment and Tourism (MET)
			  ✓ NGOs & Private sector

Producer	Status	Official mapping agency	▪ Official producers
			
			  ✓ GRN (e.g. Min. Lands: topographic maps, cadastral; Min. Mines & Energy (MME): geotopo; Min. Works & Transport (MWT): , Min. Agric: soil, vegetation, boreholes; MET: EIA, EIS); and parastatals (e.g. NSA: stats)
			
			    • satellite data via SADC AMESD project
			
			  ✓ Local authorities: cadastral, services etc.

		Commercial mapping agency	None yet
		Community interest	None yet
		Crowd source	Environmental Information System (EIS) managed by Raison
	Motivation		
		Special interest	None yet

		Economic	▪ Profit - for private companies
			▪ Mandate (not for profit)
			  ✓ All official producers for socio-economic development
			  ✓ Organization and employees are paid to do it

		Process	None yet
	Role		
		Captor of raw data	Official data producers
		Submitter of revision notice	▪ Currently: Users notify producers about errors; producer takes decision regarding rectification. ▪ In future: Secretariat will receive notices, and take decision regarding rectification.
		Passive producer	None yet
		Database administrator	▪ Each data producer ▪ NSA CSD Secretariat in future
	Skill	Neophyte	None yet
		Interested amateur	None yet
		Expert amateur	None yet
		Expert professional	Yes, including pseudo professionals
		Expert authority	Yes, contract based
Provider			
	Data provider	A producer who provides own data	▪ All official producers ▪ ConInfo
		Data distributor	▪ NSA ▪ EIS (crowdsourcing) ▪ MET (e.g. Nam. Atlas and Northern Environmental Profile and Caprivi Environmental Profile) ▪ Fire, etc (EMIN2)

			Data arbiter	In Future: NamSDI Secretariat within NSA
	Service provider	Producer is its own service provider		Yes, EIS
		Service distributor		None yet
		Service arbiter		None yet
Broker				
	Crowd-sourcing facilitator			EIS
	Finder	Clients/users finder		Private companies (PCs) (e.g. GeoCarta, Geobusiness Solutions, Prime GIS Technologies (PGT)
		Providers finder		PCs (e.g. PGT)
	Harvester			In future: NamSDI Secretariat
	Cataloguer			EIS, NamSDI Secretariat
	Négociant			PCs (e.g. PGT), Technical advisors in Ministries, NGOs.
Value-added reseller (VAR)				
	Publisher			▪ Satellite receiving station (Polytechnic of Namibia, SADC, etc – image processing), Okaukuejo (Etosha) satellite imagery o Value: (image processing), Clark Labs (NDVI for Namibia), ▪ Ministry of Agriculture (value-addition: fire monitoring and alerts) early warning ▪ Ministry of Lands: publish topographic maps with data from different sources (e.g. schools)

			■ NSA & Geological Surveys: maps
	Aggregator/ integrator	Service integrator	None yet
		Data and metadata aggregator/ integrator	MET (Nam. Atlas), NSA (Poverty Atlas), MLR (Land-use suitability; regional environmental impact assessments)
End user			
	Naive consumer		■ Some citizens and visitors, ■ Some government employees, ■ Some consultants and private companies
	Advanced user		■ Some citizens and visitors, ■ Some government employees ■ Some consultants and private companies

NamSDI secretariat within NSA is mandated *secretarial* roles and responsibilities. A number of individuals working in various organizations such as the NPC (before the Act), NSA (after the Act), GRN Ministries, Non-governmental organizations and the private sector expressed the need for standardized data and are working as NamSDI *champions*.

4.2 Producer

Producers of official geospatial datasets for the NamSDI community include GRN ministries and their agencies (Table 2) because of their budgets, resources, expertise, mandates and interests. These datasets include topographical and cadastral maps (Ministry of Lands), hydrographic and hydrological (Min. Agriculture), meteorological and transportation (Min. Works), geological (Min. Mines and Energy), social statistical (NSA), environmental (Min. Environment) and other fundamental datasets produced by other mapping agencies. Local authorities also produce spatial data such as cadastral and infrastructure services. No *commercial mapping agencies* or *community interest* groups produce spatial data suitable for NamSDI at present. Meanwhile, the environmental information system (EIS) managed by Raison (private company) produces *crowd sourced* geospatial datasets from citizens.

4.3 Provider

The majority of official producers and parastatals provide data or services for their internal use and for others through NamSDI. ConInfo, established and managed by the Ministry of Environment, is one such example. The current NamSDI legislation mandates the NamSDI secretariat within NSA to establish the metadata catalogue of all official data producers and played the role of *data distributor* (under NPC, before the Act). In addition to EIS mentioned earlier, the Ministry of Environment and Tourism is another example of web-based distributors of spatial data in Namibia. The NamSDI Secretariat within NSA will in future play the role of *data arbiter* (Table 2).

4.4 Broker

Crowd sourcing facilitators for NamSDI include the EIS, which brings together End Users and Providers and allow groups of individuals to access on-demand data and services. Finders of spatial data and services through NamSDI involve private companies as *client/user finders* (Table 2) as they promote and sell a portfolio of data and services from *producers, providers and VARs*, to end users. Many of the private companies work also as *provider's finders* by sourcing data or services for NamSDI, through mandates and contracts.

NamSDI secretariat through the act is mandated to play the roles of *harvester* and *cataloguer* of geospatial metadata and services, by building and maintaining the catalogue for NamSDI. Private companies, technical advisors in Ministries and NGOs are the major *négociants* who bring end users and providers together and assist in the negotiation of contracts.

4.5 Value-Added Reseller

Official mapping agencies (Table 2) and satellite receiving stations (e.g. Polytechnic, SADC image processing, Okaukuejo's Etosha satellite imagery, National remote sensing's fire monitoring) are examples of VAR *publishers* whose duties involve collecting data from various sources, and integrating and editing them to produce new products. The VAR *service integrator* is currently non-existent, and *data and metadata aggregators/integrators* include Ministry of Environment (Namibian Atlas), NSA (Poverty Atlas) and Ministry of Land (Land-use suitability; regional environmental impact assessments).

4.6 End User

End user stakeholders use NamSDI for their intended purpose and include citizens and visitors, government employees, consultants and private companies who are either *naive consumers* or *advanced users.* End users who use whatever spatial datasets available due to limited ability in discerning the quality of the data or services are categorized as naive consumers. In contrast, advanced users are experts with the ability to make informed decisions regarding the quality of geospatial data and services.

5 Discussion

As emphasized throughout this chapter, the motivation of Namibian stakeholders is to contribute towards the successful implementation of NamSDI. The results of our research show that the use of a high-level ICA model in contextualizing a policy and legal dependent NamSDI is robust. It is prevalent in NamSDI that the motivation of GRN, agencies, NGO's and private sector involved direct economic and financial rewards, business awareness promotion and end users unwilling to pay for institutional data (Cooper et al. 2012). Special interest groups such as donor agencies, academic institutions and non-governmental organizations produce data for their own specialized interest. Meanwhile, academic and research institutions also produce data as part of their training processes.

This study also reveals that many NamSDI stakeholders assume multiple roles and use cases, for example, all *official mapping agencies, commercial agencies, community interest mappers and crowd sourcers* can be classified as *captors of raw data.* The NamSDI secretariat within NSA acts as recipient of *submitted notices* in order to revise or correct data. Since NamSDI is still at an infancy stage, some of the stakeholder sub-types, like the roles of *database administrators,* are executed by each official producer.

Neophyte and *Interested amateur* producers *do not exist yet* in NamSDI. *Special interest* and rainfall data *crowd sourcing* are producers of data with *expert amateur* skills. All NamSDI spatial data producers are *expert professional* and/or *expert authority* because they studied and practice the subject and rely on that knowledge when engaged in spatial data production.

There are several challenges to consider regarding the significance of using the ICA model in describing stakeholders in NamSDI. No NamSDI official stakeholder classification or study is documented that could serve as a reference point for this chapter. Some of the ICA model stakeholders and sub-types were not applicable due to the infancy status of NamSDI (i.e. no study was conducted to assess the skills of stakeholders in their capacity as producers and service providers in NamSDI). In respect of these viewpoints, the contribution of this chapter enhances the understanding of NamSDI.

6 Conclusion

The aim of this chapter was to model the NamSDI stakeholders as types and subtypes of the ICA's model stakeholders. The experience gained in modelling the stakeholder types and subtypes improved the understanding of the NamSDI and served to test the behavior and applicability of the high level abstract model to a specific SDI instance. Exploring NamSDI stakeholder types and subtypes as ICA's model stakeholder revealed an interesting trend of current research in defining and characterizing stakeholders of a policy and legalistic dependent SDI.

The ICA's model helped in identifying and contextualizing NamSDI stakeholders, as well as their roles and responsibilities. Thus, NamSDI stakeholder modelling has produced results that make intuitive sense from an SDI perspective and has enhanced the ability to define stakeholder type and subtype trends of legislative dependent SDIs in a relatively short period. It is also important to note that NamSDI stakeholder identification cannot be exclusively restricted to a stakeholder model but should include SDI objectives, purpose and user needs.

References

Cooper AK, Coetzee S, Rapant P, Laurent D, Peled A, Danko DM, Moellering H, Düren U (2013a) Exploring the impact of a spatial data infrastructure on value-added resellers and vice versa. (accepted for the 26th International Cartographic Conference (ICC 2013), Dresden, Germany)

Cooper AK, Moellering H, Hjelmager J, Rapant P, Delgado T, Laurent D, Coetzee S, Danko DM, Düren U, Iwaniak A, Brodeur J, Abad P, Huet M, Rajabifard A (2013b) A spatial data infrastructure model from the computational viewpoint. Int J Geogr Info Sci (in press)

Cooper AK, Rapant P, Hjelmager J, Laurent D, Iwaniak A, Coetzee S, Moellering H, Düren U (2011) Extending the formal model of a spatial data infrastructure to include volunteered geographical information. 25th international cartographic conference (ICC), Paris, 4–8 July 2011. http://researchspace.csir.co.za/dspace/handle/10204/5212

Cooper AK, Hjelmager J, Nielsen A, Rapant P (2003) Description of spatial data infrastructures (SDIs) using the Unified Modeling Language (UML). International Cartographic Association (ICA). http://researchspace.csir.co.za/dspace/handle/10204/1776

Cooper AK, Coetzee S, Kourie DG (2012) Assessing the qualities of repositories of volunteered geographical information. GISSA Ukubuzana 2012 Conference, Ekurhuleni, South Africa, 2-4 Oct 2012.

Hjelmager J, Moellering H, Cooper AK, Delgado T, Rajabifard A, Rapant P, Danko D, Huet M, Laurent D, Aalders HJGL, Iwaniak A, Abad P, Düren U, Martynen-ko A (2008) An initial formal model for spatial data infrastructures. Int J Geogr Info Sci 22(11&12):1295–1309

ISO/IEC 10746-1:1998, Information technology—open distributed processing—reference model: overview. International Organization for Standardization (ISO), Geneva, Switzerland

ISO/IEC 19501:2005, Information technology—open distributed processing—unified modelling language (UML) Version 1.4.2. International Organization for Standardization (ISO), Geneva, Switzerland

Namibian Association of CBNRM Support Organisations (2012) ConInfo. http://www.nacso.org.na/coninfo.php. Accessed 10 Oct 2012

National Planning Commission of Namibia (2011a) Statistics Act, Act No 9 of 2011, Government Notice, No. 148, Windhoek, Namibia

National Planning Commission of Namibia (2011b) Draft National Spatial Data Infrastructure Policy of Namibia, Windhoek, Namibia

Namibia Statistical Agency (NSA) (2011) National Spatial Data Infrastructure Standards Schedule of Namibia, Government Notice, No. 2011, Windhoek, Namibia

Nebert D (2004) Developing spatial data infrastructures: the SDI cookbook. http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf. Accessed 23 Aug 2009

Raison (2012) Environmental information service of Namibia. http://www.the-eis.com. Accessed 11 Oct 2012

Raymond K (n.d.) Reference model of open distributed processing (RM-ODP): introduction. Centre for Information Technology Research, University of Queensland, Brisbane 4072, Australia

Sinvula KM, Coetzee S, Cooper AK, Hipondoka M (2012) Exploring the potential suitability of an SDI model in context of the National Spatial Data Infrastructure (NamSDI) of Namibia. GISSA Ukubuzana 2012 Conference, Ekurhuleni, South Africa, 2–4 Oct 2012

Exploring the Impact of a Spatial Data Infrastructure on Value-Added Resellers and Vice Versa

Antony K. Cooper, Serena Coetzee, Petr Rapant, Dominique Laurent, David M. Danko, Adam Iwaniak, Ammatzia Peled, Harold Moellering and Ulrich Düren

Abstract A spatial data infrastructure (SDI) is an evolving concept for facilitating, coordinating and monitoring the exchange and sharing of geospatial data and services. In earlier work, we developed a formal model for an SDI from the Enterprise, Information and Computational Viewpoints of the Reference Model for Open Distributed Processing. Within the Enterprise Viewpoint, we identified six stakeholders, including a *Value-added Reseller (VAR)*, a stakeholder who adds value to an existing product or group of products, and then makes it available as a new product. A VAR is particularly important because they extend the usefulness

A. K. Cooper (✉)
Built Environment, CSIR, PO Box 395Pretoria 0001, South Africa
e-mail: acooper@csir.co.za

A. K. Cooper · S. Coetzee
Centre for Geoinformation Science, University of Pretoria, Pretoria 0002, South Africa

P. Rapant
IT4Innovations, and Institute of Geoinfomatics, VSB-Technical University of Ostrava, 17. Listopadu 15, Ostrava-Poruba, Czech Republic

D. Laurent
Institut Géographique National, 4 Rue Pasteur 94165 Saint Mandé, France

D. M. Danko
Esri, 8615 Westwood Center Drive, Vienna, VA 22182-2214, USA

A. Iwaniak
Institute of Geodesy and Geoinformatics, Wrocław University of Environmental and Life Sciences, Wrocław, Poland

A. Peled
Department of Geography and Environmental Studies, University of Haifa, 31905 Haifa, Israel

H. Moellering
Department of Geography, Ohio State University, Columbus, OH 43210, USA

U. Düren
Bezirksregierung Köln, Muffendorfer Strasse 19-21 53177 Bonn, Germany

of SDI products: high quality and useful VAR products help ensure continued funding by governments of publicly provided data. We engaged with various types of VAR around the world, to understand what encourages or inhibits VARs in an SDI, and the contributions VARs can make to an SDI. The results are described here.

Keywords Spatial data infrastructure (SDI) · Value-added reseller (VAR) · Public sector geographical information (PSGI)

1 Introduction

A *spatial data infrastructure (SDI)* is an evolving concept for facilitating, coordinating and monitoring the exchange and sharing of geospatial data and services, and the metadata about both. It encompasses stakeholders from different levels and disciplines. An SDI is more than just the technology of a geographical information system (GIS): it is a collection of technologies, policies and institutional arrangements and provides the basis for the discovery, evaluation and application of geospatial data and services (Cooper et al. 2011, adapted from Hjelmager et al. 2008 and Nebert 2004).

One SDI can be part of another SDI, either functionally (e.g.: a water SDI within a general SDI) or hierarchically (e.g.: the Europe-wide SDI, INSPIRE, which is based on the national SDIs of Member States (European Parliament 2007)). The Commission on Geoinformation Infrastructures and Standards of the International Cartographic Association (ICA) has used the *Reference Model for Open Distributed Processing (RM-ODP)* (ISO/IEC 10746-1:1998) and the *Unified Modelling Language (UML)* (ISO/IEC 19501:2005) to develop formal models of an SDI. The Commission has described an SDI from the *Enterprise* and *Information Viewpoints* of RM ODP (Hjelmager et al. 2005, 2008), and the *Computational Viewpoint* (Cooper et al. 2007, 2009, 2013).

Within the Enterprise Viewpoint, Hjelmager et al. (2008) identified six general roles of stakeholders in and around an SDI concerning the *products* of the SDI (data and services): *Policy Maker*, *Producer*, *Provider*, *Broker*, *Value-added Reseller (VAR)* and *End User*. These are not only the roles already performed within and around an SDI, but also those that should be performed. Cooper et al. (2011) took this further, identifying various special cases of these general roles (termed 'sub-types' in that paper). Any one individual or organisation can perform or fulfill several general or special roles.

Cooper et al. (2011) described a *Value-added Reseller (VAR)* as a general role which adds some new feature to an existing product or group of products, and then makes it available as a new product. They described a *Broker* as a general role which brings End Users and Providers together and assists in negotiating contracts between them. The Broker and the VAR are at the interface between the SDI and

the End Users and are the bridge between the Producers and/or Providers and the End Users.

The key difference is that the VAR anticipates what the market will need, adding value to what is available to supply data and/or services; while the Broker assesses the supply and demand and exploits the opportunities of bringing them together. Obviously, one organisation can be both a Broker and a VAR (and a Producer and a Provider, etc.). We considered only the VAR, because their offerings are more immediate, available off the shelf and meet broader needs, and because we anticipate there are more VARs than Brokers. We are considering assessing the impact of Brokers.

Cooper et al. (2011) identified the special cases of a VAR (see Fig. 1):
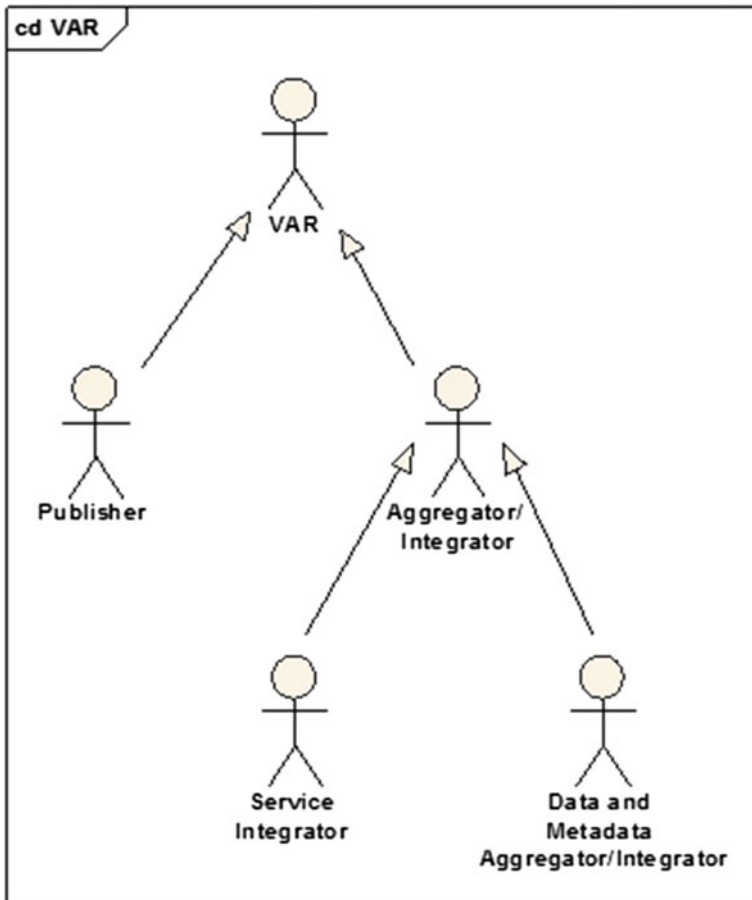
Fig. 1 Specializations of the VAR (Cooper et al. 2011)

- *Publisher*—takes data from various sources and cleans, integrates and edits them to produce a new product, such as an atlas, a location-based service or adding attributes to administrative areas. A Publisher could add some of their own data. This is probably the most common function of a VAR, particularly for taking data from organisations with contiguous jurisdictions (e.g.: cadastral data from municipalities) and cleaning and integrating the data.
- *Aggregator/Integrator*
  - *Service Integrator*—chains services together, often residing in the cloud. An example would be an organisation that combines open-source web services into a new service.
  - *Data and Metadata Aggregator/Integrator*—selects, edits, enhances and combines data into a new offering:
- Conflating data sets (selecting the "best" versions of features and attributes). This differs from the *Publisher* as the *Aggregator/Integrator* will select from "competing" versions of the same data set;
- Aggregating metadata (combining metadata together, such as to provide a catalogue service, often including cleaning the metadata and converting it to conform to a standard); and/or
- Selecting and integrating different data sets and their metadata from different Providers to produce new data.

As shown in Fig. 1 (which uses UML conventions), the Publisher and Aggregator/Integrator are *specializations* of the VAR, and the Service Integrator and Data and Metadata Aggregator/Integrator are *specializations* of the Aggregator/Integrator.

A VAR does not necessarily sell its products, but could generate income from other related sources (e.g.: support services or donor funding), use its offerings to promote more valuable offerings, or even provide the products as a community service (e.g.: as a by-product of a funded activity). Often, a VAR creates the value-added data set or service for its own purposes, and then discovers a wider demand for the offering. A VAR will generally, but not always, be in the private sector.

VARs are particularly important as they fill the gap in the market between the Producers and Providers on the one hand, and the End Users on the other. They extend the usefulness of an SDI because their high quality and useful VAR products derived from the SDI help ensure continued funding by governments of PSI. The United States of America came to dominate the market for GIS during the 1970s because of the ready availability of raw data from all levels of government and of earth-observation data, that needed intermediaries (i.e.: the VARs) to clean, integrate and/or analyse the data and make them useful to the End Users.

In a country with no significant SDI up and running yet (e.g.: South Africa), there is likely to be a variety of *Publishers* filling the gap caused by the lack of an SDI. Establishing the SDI would not drive the Publishers out of business, but would allow them to move up the value chain. Similarly, such a country probably

has *Data Aggregators/Integrators*, particularly for data sets from multiple authorities without one taking overall responsibility.

Lopez-Pellicer et al. (2011) claim that orchestrated web processing services for geographical information are still used mainly at universities and research institutes, and not yet in practice. Hence, *Service Integrators* are likely to be in-house units that developed automated or semi-automated workflows to provide integrated services to their organisation. An example might be an automated imagery correction, rectification and cloud-detection workflow in a satellite imagery receiving station.

2 Background

Oxford Dictionaries (2010) defines a *VAR* as a "*value-added reseller, a company that adds extra features to products it has bought before selling them on*".

The South African VARs interviewed are aware of the newly established SDI, but the SDI is not yet delivering any products. Sebake and Coetzee (2012) analysed the motivators and barriers to address data sharing at organizations in South Africa, who agreed that a data sharing initiative would assist the public sector focus on better service delivery, while allowing the private sector to focus on developing value-added products and services. This was confirmed by the VAR responses to our questionnaire.

The European Directive on public sector information (PSI) (2003) impacted on the reuse of PSI in Europe. Dekkers et al. (2006) distinguished seven logically related dimensions for PSI reuse from the Measuring European Public Sector Information Resources (MEPSIR) study on PSI exploitation:

1. *Availability*. PSI must be available for re-use to create a market.
2. *Accessibility*. Availability leads to nothing if the PSI is not accessible.
3. *Transparency*. Accessibility needs transparency, particularly for the re-use of PSI.
4. *Accountability*. The PSI suppliers should be accountable for adhering to these conditions.
5. *Non-discrimination*. Suppliers shall apply the same conditions to all users.
6. *Actual demand*. Equal and fair access to PSI will boost the actual demand.
7. *Economic results*. Benefits both direct (more turnover for re-users) and indirect (more commercial activity based on PSI).

These dimensions were reflected in the responses from the VARs. The MEPSIR study predicted three impact typologies of the PSI Directive:

1. *Closed shop*. PSI is hardly affected by the Directive as its production is undisputedly a core public task. Often with a legal monopoly, the value chain is controlled by the public sector. Cadastral information is an example.

2. *Battlefield*. The involvement of the public sector is disputed, especially the production of value-add information. Often the public sector has advantages, such as owning infrastructure needed to generate the data. However, as technology evolves, this position is weakened. Weather information is an example.
3. *Playground*. Government either opens up, providing its resources at very lower costs or even for free, or transforms its PSI into a non-economic good.

A survey by Fornefeld et al. (2008) showed that the PSI Directive impacted on the geographical information sector, with income of geographical information re-users increasing and the market being enriched by many new re-user groups offering innovative applications. However, PSI re-users still complained about pricing and licensing.

Welle Donker et al. (2010) claim that after price, inconsistent and opaque licence conditions for public sector geographical information (PSGI) are still major obstacles for PSGI reuse by VARs. They point out the restrictions of Creative Commons (CC) licenses and introduced Geo Shared licenses, inspired by the CC concept. Both licence categories provide a way to review and categorise current licences.

Recently, the European Union began the project *smeSpire* (Smits et al. 2012) to assess the market for 'geoICT' companies, particularly small and medium-sized enterprises (SMEs), and characterize the obstacles. As many VARs are SMEs, our research should inform smeSpire and similar initiatives.

3 Methodology

We identified several different types of VARs around the world with whom we are familiar, to try to understand whether or not an SDI encourages or inhibits VARs, and the contributions that VARs can make to an SDI. We asked them a set of questions through telephonic or face-to-face interviews, or via email. Specifically, we interviewed two large, multinational private-sector VARs (one in two different countries), seven medium or small private sector VARs and one public sector VAR, in the United States of America, South Africa, Czech Republic, Poland, France and Israel. The questionnaire asked about their VAR activities and offerings, their understanding of SDIs, and whether laws, licences, costs, metadata and other issues encourage or inhibit them.

As this was exploratory, qualitative research and as we aimed at questions that did not influence the other answers, we sent the questionnaire to the respondents in advance, so they could prepare, as appropriate. We also did not try to obtain a representative sample or eliminate 'duplicates' from different countries: by covering different countries we had a variety of legislative, business and cultural contexts in which the VARs operate. Clearly, such a qualitative study cannot produce any statistics, but it has identified key issues regarding the successful creation and operation of an SDI, and also for enabling VARs to provide more interesting products.

4 Results

The following is a summary of the responses we received from the VARs, who all agreed they are VARs, but perhaps only at the end of a long chain of processes (though one does not sell data):

1. The main VAR activities are linking, adapting, updating, geocoding, consolidating and integrating data bases and data records (both base and specialist data, from both public and private sources), developing new products, sometimes incorporating their own data, analysis and forecasts, and advising clients. They generally deliver their products through web services and geoportals. The VAR will often also provide service-level agreements (SLAs) for data access and response times. As one VAR put it: they are a throat to grab! The VAR is willing and able to deal with queries and other support issues that might overwhelm the original Producers or Providers (especially under-resourced municipalities) and hence discourage the Producer or Provider from publishing data.

2. Some VARs were familiar with the national SDI activity in their countries (whether or not it is delivering any products now), even helping build them, but others were not. This emphasizes that the stakeholders in an SDI should not assume that the SDI, its activities and benefits are well understood in the community or considered a priority yet—particularly by SMEs.

3. SDIs provide services, metadata and base data, sometimes using standards such as Web Map Service (WMS) and Web Feature Service (WFS). They also develop standards. SDIs can encourage the development of crowd-sourced or user-generated applications, such as vehicle navigation.

4. Some laws encourage VARs, especially the promotion of access to public information and in providing a framework for cooperation between data custodians. This stimulates VAR activities. However, some laws inhibit VARs, particularly licencing conditions preventing selling the value-add, as Welle Donker et al. (2010) found. Even where new laws encourage PSGI reuse, they might not be well known and the old habits of hoarding data can prevail. The situation does vary across countries and cultures. There can also be contradictions in legislation. Delays in implementing legislation inhibit investments. There are uncertainties over issues such as liability and privacy, which can discourage sharing PSGI, but good laws on these issues can create trust. Registration of professionals (not only those employed by VARs) will improve consumer confidence.

5. Standards are essential for integrating data sets, particularly across jurisdiction boundaries. They reduce costs and provide competition for products. However, better training on standards is required and there are important aspects that are not yet standardized.

6. Capturing data at the finest spatial level makes the data less vulnerable to changing administrative boundaries.

7. SDIs such as INSPIRE and global trends have promoted the availability of free data, particularly to public bodies: while it takes some markets from VARs, it enables many more GIS projects. However, some PSGI is still too expensive, encouraging user-generated alternatives. Pricing should be for high volume and low margin, not low volume and high margin, as this stimulates the market. Tools are needed for assessing the costs of different options. If VARs have to pay for data, this is a barrier, particularly for SMEs that provide niche products. However, if PSGI is free, End Users might not realise how expensive it is to produce! Some private-sector VARs distribute PSGI for free, but then charge for customizing the data.

8. Some VARs spend much time negotiating license agreements with PSGI providers and licences can be unattractive and complex. One provider changed its license agreement a decade after the VAR had been integrating and reselling its datasets, threatening the VAR's investment. Having guiding principles stated clearly in legislation, will simplify value-added data and services business. Short-term rental agreements should be considered. The responses show that SDIs improve the situation.

9. Some SDIs have made licencing of PSGI easier, cheaper and with fewer restrictions. This helps VARs sell other products and services.

10. Poor quality PGSI creates a market for VARs, but also creates problems and extra work for VARs, even preventing value-add. Different applications require different levels of quality, particularly positional accuracy, currency and completeness. Quality value-add products can also provide the justification for the SDI.

11. SDIs improve data and access—even before the SDI is fully functional, because they encourage sharing. The general maturity of the market also improves data, but SDIs need to harmonize data. However, SDIs, Producers and Providers can flounder when they lose key staff.

12. Some VARs think the metadata standards require too much detail, while others would like more metadata and at a finer resolution. Standards facilitate interoperability and help prevent technology lock-in, but some users are reluctant to adopt local standards because of the perceived difficulties. Adopting standards is beneficial in the long term but painful in the short term. Adherence to standards can encourage users because of the perceived quality, and thus Producers.

13. Generally, the VARs are using cloud computing and some also resell cloud computing services. Some are experimenting with services using linked data, but others are not aware of linked data at all.

14. People change slowly. They (still) need to be reminded again and again about the benefits of geospatial data and the sharing of it.

15. Some VARs recognize an interdependent relationship between VARs and SDIs: VAR business increases if there is an SDI and SDIs depend on the success of VARs. VARs want to help SDIs and get more public bodies participating, as they are markets for them.

5 Conclusions

We report here on exploratory research on the impact of SDIs on VARs, and vice versa. Key insights include that SDIs are not well understood by VARs, the products of VARs can provide justification for funding an SDI, SDIs seem to improve data access, having multiple VARs reduces data availability risks and VARs often fulfil other stakeholder roles, especially being Brokers. VARs create jobs, resulting in more tax payers: the VAR and its (additional) employees now pay more taxes, which can be used to improve PSGI. The results confirm the importance of creating awareness of the value of spatial information and SDIs in the broader community (Rautenbach et al. 2012). This needs to be investigated further, such as through a quantitative study.

References

Cooper AK, Moellering H, Delgado T, Düren U, Hjelmager J, Huet M, Rapant P, Rajabifard A, Laurent D, Iwaniak A, Abad P, Martynenko A (August 2007) An initial model for the computation viewpoint of a spatial data infrastructure. Proceedings of the 23rd international cartographic conference, Moscow

Cooper AK, Moellering H, Hjelmager J, Rapant P, Delgado T, Laurent D, Coetzee S, Danko DM, Düren U, Iwaniak A, Brodeur J, Abad P, Huet M, Rajabifard A (2013) A spatial data infrastructure model from the computational viewpoint. Int J Geogr Info Sci 27(6):1133–1151. doi:10.1080/13658816.2012.741239

Cooper AK, Moellering H, Hjelmager J, Rapant P, Laurent D, Abad P, Danko D (2009) Detailed services in a spatial data infrastructure from the computation viewpoint. Proceedings of the 24th international cartographic conference, Santiago

Cooper AK, Rapant P, Hjelmager J, Laurent D, Iwaniak A, Coetzee S, Moellering H, Düren U (2011) Extending the formal model of a spatial data infrastructure to include volunteered geographical information. 25th international cartographic conference, Paris

Dekkers M, Polman F, te Velde R, de Vries M (2006) Measuring European Public Sector Information Resources (MEPSIR). Final Report of Study on Exploitation of public sector information—benchmarking of EU framework conditions. European Commission

European Parliament (2003) Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. OJ L 345/90. http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF. Accessed 9 March 2009

European Parliament (2007) Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an infrastructure for spatial information in the European Community (INSPIRE), OJ L 108/1. http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF. Accessed 9 March 2009

Fornefeld M, Boele-Keimer G, Recher S and Fanning M (2008) Assessment of the re-use of public sector information (PSI) in the geographical information. Meteorological Information and Legal Information Sectors. Final report. MICUS Management Consulting GmbH

Hjelmager J, Delgado T, Moellering H, Cooper AK, Danko D, Huet M, Aalders HJGL, Martynenko, A (2005) Developing a modelling for the spatial data infrastructure. 22nd International Cartographic Conference, A Coruna, Spain

Hjelmager J, Moellering H, Cooper AK, Delgado T, Rajabifard A, Rapant P, Danko D, Huet M, Laurent D, Aalders HJGL, Iwaniak A, Abad P, Düren U, Martynenko A (2008) An initial formal model for spatial data infrastructures. Int J Geogr Info Sci 22(11&12):1295–1309

ISO/IEC 10746-1:1998, Information technology—open distributed processing—reference model: overview, International Organization for Standardization (ISO), Geneva

ISO/IEC 19501:2005, Information technology—open distributed processing—unified modeling language (UML) Version 1.4.2, International Organization for Standardization (ISO), Geneva

Lopez-Pellicer FJ, Rentería-Agualimpia W, Béjar R, Muro-Medrano PR, Zarazaga-Soria FJ (2011) Availability of the OGC geoprocessing standard: March 2011 reality check. Comput Geosci 47(2012):13–19

Nebert, DD (2004) Developing spatial data infrastructures: The SDI cookbook. http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf. Accessed 23 Aug 2009

Oxford Dictionaries (2010) Oxford dictionaries. Oxford University Press. http://oxforddictionaries.com/definition/english/VAR. Accessed Oct 15 2012

Rautenbach V, Coetzee S, Smit J, du Plessis H, Muzondo I (2012) Identifying the target audiences, media and messages for SDI education and training in South Africa. GISSA Ukubuzana 2012, Ekurhuleni, 2–4 Oct 2012

Sebake MD, Coetzee S (2012) Results of three case studies for assessing motivators and barriers of address data sharing in South Africa. South Afr J Geomat 1(1):32–43

Smits P, Craglia M, Cipriano P (2012) smeSpire. http://www.slideshare.net/smespire/smespire-project-overview/. Accessed 25 Oct 2012

Welle Donker F, Van Loenen B, Zevenbergen J (2010) Geo Shared licences: a base for better access to public sector geoinformation for value-added resellers in Europe. Environ Plan B Plan Des 37(1):326–343

Part VIII
Use and Usability

Geospatial Data Collection/Use in Disaster Response: A United States Nationwide Survey of State Agencies

Michael E. Hodgson, Sarah E. Battersby, Bruce A. Davis, Shufan Liu and Leanne Sulewski

Abstract In the United States presidential disaster declarations are typically issued after major disaster events to provide assistance (in the form of monies, staff, geospatial data, etc.) to states when the disaster overwhelms the resources of the state. Geospatial support is one of the forms of assistance and a frequent item noted by Federal agencies in demonstrating their relevance. During the disaster the state is 'in charge' of the disaster response while the Federal government provides assistance. Are the geospatial data (including remotely sensed imagery of all types) needs met by the states (based on their experience)? What are the expectations of the states for Federal help in geospatial data? Are states embracing newer paradigms for collecting/exploiting geospatial data, such as volunteered geographic information or crowd-sourced data/information? In the winter of 2011–2012 a nationwide survey of the geospatial data, methods, and problems in all fifty United States emergency management offices (EMOs) was conducted. Responses to the key questions on geospatial data priorities, remotely sensed imagery, timeliness, expectations, staffing, and emerging technologies are presented in this article. This nationwide survey of state EMOs provides a unique view of the EMO director's view of geospatial methods during emergency response/recovery.

M. E. Hodgson (✉) · S. E. Battersby · S. Liu · L. Sulewski
Department of Geography, University of South Carolina, Columbia, USA
e-mail: hodgsonm@sc.edu

B. A. Davis
Science and Technology Directorate, Department of Homeland Security, Washington DC, USA

1 Introduction

The hazard cycle is thought of, for a given place, as a sequence of phases—warning/preparation, the event, response, recovery, planning/mitigation, and back to the warning/preparation stage again. Geographic information systems (GIS) and remote sensing technologies have been used in hazard-related applications for some two to three decades (Jensen and Hodgson 2006). In the mapping sciences community the prevailing assumption is that the use of GIS, GPS, and remote sensing are critical components in all phases of the hazard cycle. The frequent, but misleading, assumption is the state emergency management office (EMO) has the geospatial capabilities to manage major disasters. In early 2005, 77 % of the state-level EMOs, for instance, had at least one full-time staff that was qualified with GIS capabilities (Hodgson et al. 2010). But 23 % of the states had no GIS or remote sensing educated staff. In a disaster event whom do these states depend on for GIScience support?

The Federal government's primary coordination of geospatial support for disasters comes from the Federal Emergency Management Agency (FEMA) within the Department of Homeland Security (DHS). In response to the growing importance of geospatial technologies in the coordination of incident response and recovery operations, FEMA has created a Geospatial Information Officer to coordinate FEMA's geospatial support to state agencies during a disaster. Prior to this new FEMA coordination position, the National Geospatial Intelligence Agency (NGA), in cooperation with other Federal agencies, created a geographic database called Homeland Security Infrastructure Program (HSIP) (Department of Homeland Security 2013). HSIP contains more than 311 layers in a confidential form (HSIP Gold), primarily for situational awareness in Federal Agencies. The public release form of HSIP Gold is referred to as HSIP Freedom. Do these DHS data meet state and local needs? What geographic data are required at the state and local levels? Where/how are these data obtained and maintained? Are these data adequate from the manager's perspective?

Remotely sensed imagery, such as from airborne or satellite sensors, is widely used by public media during disaster events, and presumably by state/local responders. For example, months after the 2005 survey (Hodgson et al. 2010) no less than six separate aerial image missions were conducted for the impact area of Hurricane Katrina. One mission collected imagery over the entire Mississippi coast within the first day after Hurricane Katrina while the other aerial missions required from 7 to 10 days for image collection. The most utilized Federal response to the recent hurricane Sandy in the northeastern United States included some satellite image collections but also over 158,000 oblique images taken from low-altitude aircraft. Are the needs of the disaster managers met with these image collections? Was the imagery collected and exploited in a timely manner? When is remotely sensed imagery collection too late for disaster response/recovery? The state-EMO directors surveyed in 2005 indicated they need information on most anthropogenic related phenomena (e.g. buildings, lifelines) within 3 days of the disaster event

(Hodgson et al. 2010). Could imagery collected a week after the disaster event (as was many collections after Katrina) be useful?

In 2005, the use of GIS was considered "very important" to 88 % of the states while only 40 % said the same for remote sensing (Hodgson et al. 2010). When asked why states do not rely on remote sensing imagery more, the dominant answers were cost and collection time by over 60 % of the states.

The only nationwide (United States) survey of state agencies on the use of GIS, remote sensing, and alternative spatial technologies or spatial data needs of state-level emergency management agency's was conducted in the Spring of 2005 (Hodgson et al. 2010). Assessments for GIScience approaches in a specific hazard event have occasionally been conducted (NOAA 2001; Huyck and Adams 2002; Prechtel 2005; Parrish et al. 2007; DeCapua 2007) but they only address the one state that "hosted" the single event. Other surveys target samples of county managers on a specific topic, such as the National incident Management System (Jensen 2011). Since 2005, substantial change in the availability of technology (e.g., upgrades to GIS and remote sensing software, introduction of the Google Map API, volunteered geographic information, and use of crowd-sourced, etc.) has been made and, presumably, may have altered the emergency response mapping needs of state agencies. Other surveys have been conducted on related aspects of the disaster response phase, such as the incident Command System (ICS) and the National Incident Management System (NIMS) (Jensen and Yoon 2011).

This research reports on the geospatial data use in state-level EMOs, derived from a nationwide survey. This survey provides independent (i.e. not a survey by a Federal agency) information on the state of GIScience use, particularly remote sensing, in the state-level emergency response centers within the United States. Most importantly, the perceived impediments to the use of remote sensing, the quality of geospatial data layers, the spatial/temporal requirements for obtaining these data after the event, and the adoption of emerging technologies were probed.

The goal of this research was to identify the geospatial data needs, methods, and administrative process used by the state-level emergency management offices during a major disaster. While the full survey contained over 62 questions (45 primary questions with 17 sub-questions) this research focuses primarily on those survey questions that assist in the understanding of remote sensing and emerging technology use (or lack of) during disaster response. These questions focused on the emergency response/recovery phases are:

- What are the **priority baseline** (collected prior to a disaster) and disaster **impact** data layers needed and where do the states obtain such data?
- Are the states aware of and use the **Federal supplied data streams**, such as HSIP?
- When is remotely sensed data **too late** to be useful?
- What is the state of GIS/Remote sensing **staff**? Has this changed since 2005?
- Are states adopting **new technologies** or approaches in data collection (e.g. crowd-sourcing, VGI)?

Our survey population was the directors of the emergency management office within each state. We were interested in the perception/cognition of past, current, and future use of geospatial technology by those individuals who lead emergency response efforts within their governmental entity. We did not want to assemble multiple responses within each governmental entity and interpret for a nationwide representation (the logistics and interpretation problems of such an effort are demonstrable). We used the directors of each state office, or where they chose to designate other staff (e.g. GIS directors) to complete the survey, as the best indicators of geospatial use.

Clearly, the experience level of emergency operations directors may vary based on their personal tenure at such positions, their interaction with other governmental entities, and their experience with previous natural or technological disasters. The frequency and type (e.g. earthquake, flood, and hurricane) of natural disasters varies across the United States (Fig. 1). As noted later in this article, 96 % of the states indicated they had experienced a presidential disaster declaration (PDD) in the last 5 years.
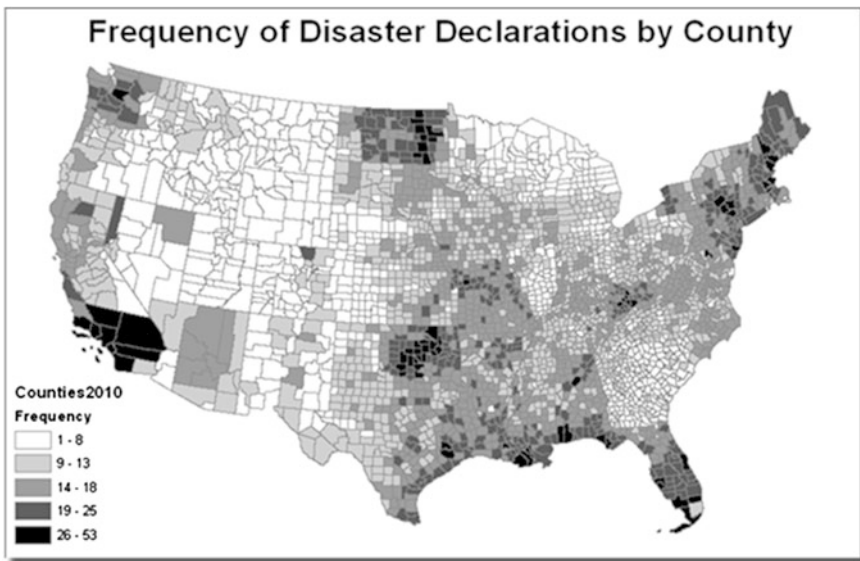
Fig. 1 Frequency of presidential disaster declarations by county since December 1964

2 Methodology

2.1 Survey Method

The survey of states in the United States was by invitation-only. We identified the single leader of the emergency management office in each state and allowed that individual to respond with his/her experience, perception, and expectations. It is critical that the governmental representative in charge of the disaster response provides the answers rather than allow a wide-open survey with an uncontrolled response population.

The survey was developed, implemented, and executed following the Tailored Design Method (Dillman et al. 2008). The state emergency management directors were contacted via regular mail introducing the survey and indicating that they would receive a formal invitation to participate in the survey within the next few weeks. The formal, personalized and signed hardcopy letter to each director was sent on November 9, 2011 inviting them to complete the online survey. This letter included the assigned login name and password for their state. In the event that there were any issues with logging into the survey, contact information for the principal investigators was listed. If the survey was submitted, a thank you letter was promptly sent out. Not surprisingly, only a few state directors immediately complied with our initial invitation. Following Dillman's protocol in the subsequent weeks we sent an email reminder, a second signed letter of invitation and a 2nd email reminder. Twenty-nine (29) states had completing the survey by January 1, 2012. From experience with the state survey conducted in 2005, we knew the remaining directors may only respond based on personal phone calls and explanations of the importance these results were to the Federal agencies and ultimately, to the Federal responses to future disasters. We made personal telephone calls to the remaining 21 state directors during the first 2 weeks of January 2012. Finally, by February 10, 2012 all 50 state directors or their designated staff completed the survey.

2.2 Survey Instrument

The survey instrument consisted of a preamble explaining the survey, a password protected login page, and four data/information categories: geospatial data use, data/resource sharing, emerging technologies, and confidential questions/answers. The instrument was constructed in close cooperation with cognizant DHS staff, the FEMA Remote Sensing Coordinator, a DHS staff focus-group meeting, and feedback from the Office of Science and Technology Policy (OSTP) Subcommittee on Disaster Reduction. Additionally, NASA program managers provided constructive feedback in refining the survey instrument. Also, the North American Counties Organization (NACO), and representatives from several county-level

emergency response agencies evaluated the final draft instrument and provided invaluable feedback on the question construction. To encourage honest and complete answers to question, and to abide by Institutional Review Board procedures, state respondents were all informed their individual answers to questions would be held confidential and only reported to Federal agencies or the public in summary form.

3 Findings

3.1 Disaster Experience

Based on the documented frequency of presidential disaster declarations (PDDs) (Fig. 1) all states have experienced a major disaster. In large part, PDDs are issued for counties and the pattern of such declarations clearly indicates a differential experience level both within and between states. The responses from our survey respondents indicate all but two states have experienced a PDD during the last 5 years. Of those 48 states that had experienced a major disaster, 77 % said they used either satellite or aerial imagery over the disaster area in the response/ recovery phases. 92 % of those using imagery indicated the type of imagery was airborne (only 44 % utilized satellite imagery). Clearly the states have greater access to, and may prefer, the use of airborne imagery.

How did the states obtain imagery? Fifty percent of the states relied on imagery collections by the Federal government and 72 % collected imagery (e.g. oblique photographs/images or vertical mapping imagery) from airborne platforms. 67 % of the respondents indicated the imagery was sufficient for their needs. In confidential responses, several states indicated that limitations to the use of imagery were often based on challenges of requesting imagery (if they were not able to collect independently), ability to collect imagery quickly enough for use in the response phase, and questions of how the imagery collection would be funded.

3.2 Pre-Event Geospatial Data Needs

Baseline geospatial data representing the natural and built-up environment prior to a disaster event is used by counties for management purposes but essentially only useful to the state EMOs for mitigation/planning. During and after the disaster event baseline data provides detailed information on what was "there", prior to the disaster, may be useful for change detection, and is used for positioning resources in response/recovery operations. Since hurricane Katrina in 2005 the DHS has funded a large effort to create and maintain the HSIP-Gold and Freedom geospatial databases. Presumably these data sources are of high importance to state EMOs.

We also sought to identify the primary source of baseline geospatial data the states rely on for each of the layers.

In the 2005 survey the baseline data layers of extreme importance in the response and recovery phases were transportation, population distribution, and location of structures. The transportation data theme was clearly the most important baseline data in the response/recovery phase. Oddly, the two spatial data themes that are commonly mapped with remote sensing technologies, land cover and terrain, were listed as the least important data types of any hazard phase.

Many states indicated an "other" source (65 and 76 % for buildings and parcels, respectively). We do not know for certain what this other source might be but hypothesize the other source may be a state-wide consortium where counties or local agencies within the state voluntarily contribute their parcel/building footprints.

In the 2011 survey, the most important baseline data layer was the **critical infrastructure** data layer (as indicated by 74 % of the states in Table 1). (Note: We did not use the critical infrastructure or aerial imagery data layer as choices in the 2005 survey and thus, do not have directly comparable responses). The source of the critical infrastructure layer was either in-house (38 %), other (32 %), or from HSIP (30 %). The HSIP data stream is not a primary data source for any of the baseline data layers (i.e. pre-event). In fact, the highest percentage of responses by the states was the 30 and 29 % reliance on HSIP for the critical infrastructure and energy/fuel supplies data layers, respectively. This observation raises an interesting question about the overall importance or lack of awareness of the HSIP data stream in planning for and during the disaster response phase of the hazard cycle.

Table 1 Source of **baseline** (i.e. prior to disaster event) geospatial data and percentage of states in 2011 indicating this data layer was a priority

Data type	Source of data (% of states)				Highest priority
	In-house	HSIP	Commercial vendor	Other	
Building footprints	21	0	15	65	*4*
Building/parcel characteristics	12	2	10	76	*18*
Communications networks	**36**	15	19	30	*20*
Energy and fuel supplies (e.g., electric, gas, etc.)	18	**29**	18	35	*20*
Critical infrastructure (e.g., hospitals, schools)	**38**	**30**	0	32	***74***
Land use or land cover	11	5	7	77	*4*
Hydrography	15	2	7	76	*0*
Population distribution	11	9	11	69	***33***
Sewer/water/utilities	18	11	5	66	*6*
Shelter locations	**65**	0	6	29	*22*
Elevation	11	0	11	78	*16*
Transportation networks	26	9	9	58	***39***
Aerial imagery (vertical or oblique)	17	0	29	54	***45***

The most important data layer after critical infrastructure was **aerial imagery**. Interestingly, aerial imagery is not a typical GIS data layer that may be directly used in a GIS analysis. In fact, imagery is a source of information that must be abstracted, such as through visual analysis. In hurricane Katrina and Ike imagery was a key basemap for positioning other features on cartographic products.

As in the 2005 survey, transportation and population were also selected in the 2011 survey as data layers of high priority. More than half (65 %) of the state-level EMOs maintain the data layer of **shelter locations.** All other data layers are obtained from a combination of HSIP, commercial vendors, or "other".

3.3 Disaster Impact Data

Of particular interest in this research were the geospatial data types representative of the **disaster impact** that were of the highest priority (Table 2). We listed seven data layers indicative of the built-up environment for which the respondents selected. A disaster impact data layer represents the damaged features for that theme. The responses indicate a clear set of three priority data layers. Not surprisingly, **critical infrastructure** again appeared as the data layer of highest priority with 69 % of the states (Table 2). **Disaster extent** and damage to **transportation features** was almost the same priority level. So where do states obtain data on damaged features? Do they contract aerial flyovers, airborne or satellite imagery, or ground surveys?

The use of emergency responders on the ground was the primary source of all data layers except for energy and fuel supply damage and disaster extent. The disaster extent source was from airborne flyovers, not airborne/satellite mapping imagery. In the 2011 survey, 32 % of the states indicated they had a mission responsibility to acquire satellite/aerial imagery in the next major disaster. When asked if the Federal government is expected to collect airborne/satellite imagery in

Table 2 Source of **impact** data (i.e. post-disaster event) geospatial data and percentage of states in 2011 indicating this data layer was a priority

Data type	Source of data (% of states)				Priority level
	Ground survey	Airborne flyover	Airborne or satellite imagery	Other	
Building damage	**75**	17	8	0	*31*
Communication network damage	**62**	9	0	30	*37*
Disaster extent (e.g., burned area)	28	**46**	16	10	**67**
Energy and fuel supply damage	**48**	2	0	**50**	25
Critical infrastructure (e.g., hospitals, schools)	**80**	4	4	12	**69**
Sewer/water/utilities	**70**	2	0	28	*8*
Transportation damage	**59**	27	4	10	65

the response phase 46 % of the states indicated "only if requested" while 30 % indicated "regardless of whether requested". 54 % of the states expected the Federal government would collect such imagery with "no cost" to the states.

When asked about institutional problems in acquiring/sharing geospatial data 51 % of the states indicated difficulties in working with other entities. The entity the states experienced the most problems with was with Federal agencies (73 % of states indicating this).

3.4 When Disaster Impact Data are Needed

Perhaps the most difficult aspect of using geospatial information, such as remotely sensed imagery, during a disaster event is obtaining the data/information in a timely manner. Immediately following a disaster event police/fire department personnel and other emergency responders are deployed to view the impact area and provide situational awareness. This process is referred to putting 'boots on the ground' or 'eyes in the air', depending on terrestrial or airborne vantage points. Unlike most other applications of remotely sensed data collecting airborne/satellite imagery within 24 h of the event is challenging. Few previous studies exist to indicate how valuable new remote sensing collections are after the disaster event—or 'n' days after the disaster event. Hodgson et al. (2010) surveyed the emergency management offices in the United States and found practically all (90 %) state offices needed information on building damage within 3 days after the disaster event. However, the immediacy of information varied by damage type. For instance, only 50 % of states indicated they needed information on crop/vegetation damage within 3 days of the event. A theoretical information-lag time curve was suggested to relate the relative value of remotely sensed derived information to the lag-time after the disaster event (Fig. 2). The understanding is if the information is provided after some delay other sources (e.g. 'boots on the ground') will provide the same information and thus, imagery is of less value.

In this 2011 survey, 30 % of the states indicated building damage information obtained from remotely sensed imagery after 72 h (3 days) was too late. In other words, 70 % of the states indicated critical infrastructure damage was needed within 3 days of the event. 16 % of states indicated information from imagery after 7 days was still useful. From a modeling perspective, the value of information to lag time for all damage categories examined followed a reverse exponential trend. The value of information quickly dropped with increasing lag time and then became a constant value for a lag-time of 7 days and more. The pattern for other disaster impact data layers was similar (Fig. 3).
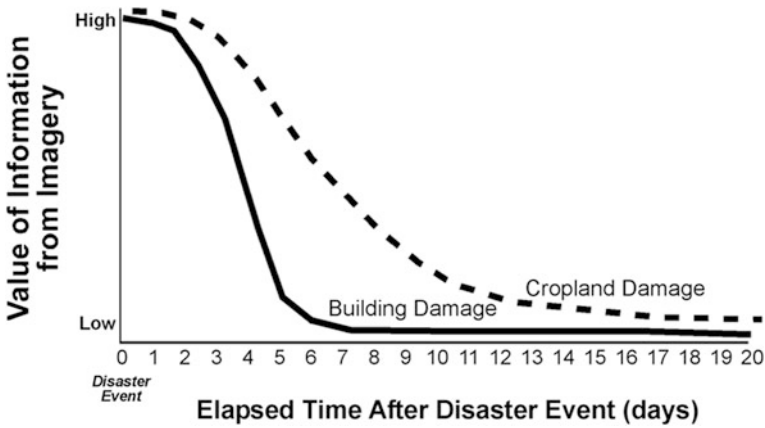
Fig. 2 Theoretical curve illustrating when post-disaster imagery information is too late for the emergency response phase (after Hodgson et al. 2010)
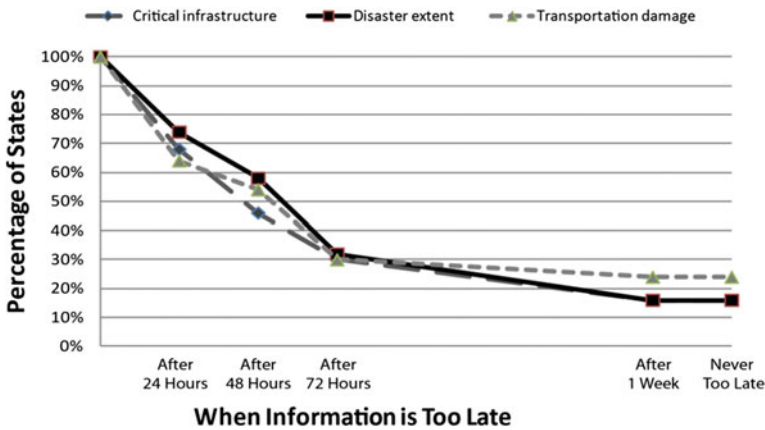
Fig. 3 Empirical curve derived from state responses for when post-disaster imagery information is too late for the emergency response phase

3.5 Staffing Needs

Adequate skilled staff is often cited as an impediment to using GIScience approaches in disaster management (Zerger and Smith 2003). In a series of questions we probed issues related to geospatial staffing (e.g. number of, expertise, etc.). For example, we obtained information on the GIS-based models (e.g. HAZUS, ALOHA, HURREVAC) used in disaster response and the limitations of using such models. By far the most frequent limitation cited was "user knowledge required" (18 % of the states). Only two states indicated "no limitations". What

we experienced in the 2005 survey of states was a deficiency in geospatial staff within the emergency management agency. For example, 23 % of the states in 2005 indicated they had no geospatial analysts within their emergency management agency. This staffing problem has largely changed. In the 2011 survey 34 % indicated they only had one such analyst. Only two states did not have at least one geospatial staff in their agency. 90 % of the states had between one and five geospatial staff members. With respect to the exploitation of remotely sensed imagery (aerial or satellite), however, only 26 states indicated they had no staff that are responsible for such imagery. In separate questions on memorandum of understanding/agreements (MOU/MOA) and assistance, the 56 % of the states indicated they rely on additional staff support through an MOU/MOA. Thus, the reliance on federal or other state/county agency support during a disaster response for imagery exploitation is still a problem.

3.6 Crowd Sourced and VGI Data

The research emphasis on volunteered geographic information (VGI) is one of the dominant paradigms in GIScience this decade (Goodchild and Glennon 2010; Heipke 2010). These methods are considered by several to be particularly useful in disaster response (Poser and Dransch 2010; Roche et al. 2011). VGI approaches are currently under investigation by FEMA, NOAA, the National Weather Service, and other agencies for acquiring near real-time information in disaster events. FEMA is investigating the potential of crowd-sourcing for exploiting image interpretation associated with hurricane Sandy in 2012. The infrastructure for utilizing VGI to understand the unfolding disaster is available although states may not have the capacity to utilize such approaches. Are state EMOs currently using or considering using such approaches? We did not separate questions on crowd-sourcing from VGI but grouped them in the survey questions. We asked the states if they were either using or considering the use of crowd sourced or volunteered geographic information (VGI) data during the response/recovery phases of the disaster. 58 % of the states indicated they have an interest in utilizing crowd-sourced and VGI data. Surprisingly, only 18 % of the states indicated they were currently using crowd-sourced/VGI data. Subsequent questions revealed states had concerns about the reliability of VGI data (78 %) or quality of VGI data (66 %). 52 % said they saw a limitation of merely setting up a crowd-sourced or VGI system.

4 Conclusions

Findings from this nationwide survey of U.S. state EMO offices provides important information on the geospatial data needs and expectations in the disaster response/recovery phases of the disaster cycle. GIScientists interested in contributing to the

solution in disaster response should focus on the priority disaster impact data layers of critical infrastructure, disaster extent, and transportation. To be largely useful, such damage information needs to be available within 72 h of the event and becomes decreasingly useful after even 24 h. There are considerable expectations of the Federal government to lead on the collection of image data. Although practically all state EMOs have a GIS analyst more than half do not have a remote sensing analyst, suggesting exploitation of imagery must be conducted by others. Finally, the state EMOs share considerable interest in using VGI or crowd-sourced approaches but have strong reservations of the reliability and quality of such derived data.

This article presented results from part of the state survey of EMO agencies conducted in 2011. A companion survey of U.S. counties was also conducted concurrently with the state survey providing an understanding of the similarities and disparities between state and county geospatial data needs, approaches, and problems. It should also be reported that this survey occurred prior to Hurricanes Isaac and Sandy and do not reflect improvements made at Federal, State, and local levels in the use of remote sensing and GIS technologies for those incidents. Furthermore, this research did not investigate the variable of disaster incident scale as a determinant of the temporal value of remote sensing data. It may be that disasters at the scale of Hurricanes Katrina and Sandy or the next big earthquake require the collection and analysis of remote sensing to address critical response questions past the stated 72 h timeline. As the use of remote sensing for disaster response improves new questions may be discovered that extend the use of airborne and satellite images into the longer recovery period. That would mean collecting the appropriate imagery to support the answers to those questions.

References

DeCapua C (2007) Applications of geospatial technology in international disasters and during Hurricane Katrina. Mississippi State University Coastal Research. http://www.gri.msstate.edu/research/katrinalessons/Documents/GeoSp_Tech_Applications.pdf. Accessed November 4, 2008

Department of Homeland Security (2013) Infrastructure protection partnerships. http://www.dhs.gov/infrastructure-information-partnerships. Accessed April 16, 2013

Dillman DA, Smyth JD, Christian LM (2008) Internet, mail, and mixed-mode surveys: the tailored design method. Wiley, Hoboken

Goodchild MF, Glennon JA (2010) Crowdsourcing geographic information for disaster response: a research frontier. Int J Digital Earth 3(3):231–241

Heipke C (2010) Crowdsourcing geospatial data. ISPRS J Photogramm Remote Sens 65:550–557

Hodgson ME, Davis BA, Kotelenska J (2010) Remote sensing and GIS data/information in the emergency response/recovery phase. In: Showalter PS, Lu Y (eds) Geotechnical contributions to urban hazard and disaster analysis. Springer-Verlag, New York, pp 327–354

Huyck CK, Adams BJ (2002) Emergency response in the wake of the world trade center attack: the remote sensing perspective, vol. 3. Engineering and organizational issues related to the World Trade Center terrorist attack. http://mceer.buffalo.edu/publications/sp_pubs/WTCReports/02-SP05-screen.pdf. Accessed November 1, 2012

Jensen JA (2011) The current NIMS implementation behavior of United States counties. J Homel Secur Emerg Manag 8(1):4

Jensen JR, Hodgson ME (2006) Remote sensing of natural and man-made hazards and disasters, Chapter 8. In: Ripple B (ed) Manual of remote sensing. American Society for Photogrammetry and Remote Sensing, Bethesda, pp 401–429

Jensen JA, Yoon DK (2011) Volunteer fire department perceptions of ICS and NIMS. J Homel Secur Emerg Manag 8(1):1547–1571

NOAA Coastal Services Center (2001) Lessons learned regarding the use of spatial data and geographic information systems (GIS) during Hurricane Floyd. U.S. National Oceanic and Atmospheric Administration, NOAA/CSC/20119-PUB, NOAA Coastal Service Center, Charleston

Parrish DR, Breen JJ, Dornan S (2007) Survey development workshop for the southeast region research initiative (SERRI) project. Mississippi State University Coastal Research and Extension Center, Gulfport

Poser K, Dransch D (2010) Volunteered geographic information for disaster management with application to rapid flood damage estimation. Geomatica 64(1):89–98

Prechtel N (2005) GIS-based activities related to the 2002 summer flood in the Dresden Area. Kartographische Bausteine—Cartographic Cutting-Edge Technology for Natural Hazard Management 30:105–116

Roche S, Propeck-Zimmermann E, Mericskay B (2011) GeoWeb and crisis management issues and perspectives of volunteered geographic information. GeoJournal 1–20

Zerger A, Smith DI (2003) Impediments to using GIS for real-time disaster decision support. Comput Environ Urban Syst 27(2):123–141

Commonalities and Differences in Eye Movement Behavior When Exploring Aerial and Terrestrial Scenes

Sebastian Pannasch, Jens R. Helmert, Bruce C. Hansen, Adam M. Larson and Lester C. Loschky

Abstract Eye movements can provide fast and precise insights into ongoing mechanisms of attention and information processing. In free exploration of natural scenes, it has repeatedly been shown that fixation durations increase over time, while saccade amplitudes decrease. This gaze behavior has been explained as a shift from ambient (global) to focal (local) processing as a means to efficiently understand different environments. In the current study, we analyzed eye movement behavior during the inspection of terrestrial and aerial views of real-world scene images. Our results show that the ambient to focal strategy is preserved across both perspectives. However, there are several perspective-related differences: For aerial views, the first fixation duration is prolonged, showing immediate processing difficulties. Furthermore, fixation durations and saccade amplitudes are longer throughout the overall time of scene exploration, showing continued difficulties that affect both processing of information and image scanning strategies. The temporal and spatial scanning of aerial views is also less similar between observers than for terrestrial scenes, suggesting an inability to use normal scanning patterns. The observed differences in eye movement behavior when inspecting terrestrial and aerial views suggest an increased processing effort for visual information that deviates from our everyday experiences.

S. Pannasch (✉) · J. R. Helmert
Department of Psychology, Engineering Psychology and Applied Cognitive Research,
Technische Universitaet Dresden, Dresden, Germany
e-mail: sebastian.pannasch@tu-dresden.de

B. C. Hansen
Department of Psychology, Neuroscience Program, Colgate University, Hamilton, NY, USA

A. M. Larson
Department of Psychology, University of Findlay, Findlay, OH, USA

L. C. Loschky
Department of Psychology, Kansas State University, Manhattan, KS, USA

1 Introduction

Under most circumstances, vision is the dominant sensory modality in humans. During visual perception, information is sampled from the environment via *active vision* (Findlay 1998). Saccades—fast ballistic movements—direct the foveal region of the eyes from one fixation point to another. During saccades, the intake and processing of visual information is largely suppressed and is therefore limited to the periods of fixations, when the eyes are relatively still. This interplay of fixations and saccades is essential, as highest visual acuity is limited to the small foveal region. Eye movement behaviour in many everyday situations, such as reading text or inspecting images, can be described as an alternation between fixations and saccades.

Fixation durations vary a great deal from one fixation to the next. It has been suggested that the length of a fixation is determined by information processing and by eye movement pre-programming (Rayner 1998). Fixation durations typically range from roughly 100 to 500 ms, but can last up to 2–3 s in some cases (Rayner 1998). Similarly, the length of saccades generally varies from between less than 1 to 130° of visual angle (Land 2004). Importantly, Velichkovsky et al. (2005) reported particular relationships in the variation of fixation durations and saccade amplitudes that were related to certain modes of visual processing. Specifically, they found fixations of shorter durations (below 180 ms) are often associated with larger saccades; this combination was termed *ambient processing*, which is assumed to serve the processing of spatial orientation and localization. Moreover, the combination of longer fixations and shorter saccade amplitudes was termed the *focal processing mode*, which is assumed to be concerned with the analysis of object features. The time course of these two processing modes has been investigated under different conditions of free viewing, which has revealed a systematic relationship: during early phases of scene inspection, the ambient mode seems to dominate, giving way to the focal processing mode with increased time (Pannasch et al. 2008). This relationship has been observed across different types of stimuli and various visual tasks. However, the corresponding analyses were relatively coarse (typically consisting of the comparison of gaze behavior in two 2-s time periods, i.e. early [0-2 s] versus late [4-6 s]).

The present experiment analyzes the time course of viewing behavior in greater detail by showing real-world scenes from different views and under different display conditions: Photos of natural scenes taken either from the terrestrial or aerial perspective were presented either upright or inverted. The research question was whether we would obtain similar gaze patterns to those reported previously when looking at upright terrestrial views. By contrasting the viewing behavior in

this particular case with images of different perspectives (aerial vs. terrestrial) and orientations (upright vs. inverted), we expected to gain further insights about the interplay between ambient and focal processing modes. Particularly, we know from gist recognition studies that compared to upright terrestrial views, both inverted terrestrial scenes and upright (as well as inverted) aerial scenes are much harder to recognize within the time course of a single fixation, and thus appear to require more than a single fixation to reach a high level of gist recognition (Loschky et al. 2010). Thus we predicted differences very early after the image onset; fixation durations should be shortest for terrestrial upright views. Starting with this hypothesis we examined whether the different viewing conditions would influence only the initial gaze behavior or gaze behavior observed throughout a longer time period of scene inspection.

2 Method

2.1 Subjects

Thirty students (19 females) at Colgate University with a mean age of 18.5 years took part in this experiment. One subject was removed from the sample because their performance in image categorization was at chance level. All subjects had normal or corrected-to-normal vision and were given course credit for their time. The study was conducted in conformity with the declaration of Helsinki, and Institutional Review Board-approved written informed consent was obtained.

2.2 Apparatus

Participants were seated in a dimly illuminated, sound-attenuated room. Eye movements were sampled monocularly at 1 kHz using the SR EyeLink 1000 infrared eye tracking system with on-line detection of saccades and fixations and a mean spatial accuracy of better than 0.5°. Saccades were identified by deflections in eye position in excess of 0.1°, with a minimum velocity of $30° \text{ s}^{-1}$ and a minimum acceleration of $8000° \text{ s}^{-2}$, maintained for at least 4 ms. Pictures were displayed using an VidiaQuadro NVS 285 dual DVI/VGA graphics card and a CRT display (21-inch Viewsonic, G225fB). Maximum luminance output of the display monitor was 100 cd/m^2, the frame rate was set to 85 Hz, and the resolution was set to 1024 × 768 pixels. The monitor was viewed at a distance of 91.5 cm. Head position was maintained with an SR Research chin and forehead rest.
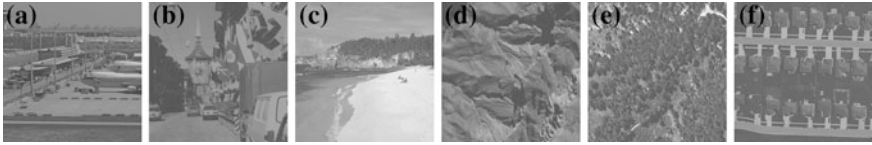
Fig. 1 Examples of the stimuli. Images (**a–c**) represent terrestrial views of the categories airport, city, and coast; images (**d–f**) show aerial views of the categories mountain, forest, and suburbs

2.3 Stimuli

Grayscale images of real-world scenes from six possible categories, viz. airports, cities, coasts, forests, mountains, or suburbs (Loschky et al. 2010), served as stimuli (Fig. 1). Each category was comprised of 60 unique images. Of these, 30 images presented an aerial view (as viewed from a satellite) and the other 30, a terrestrial view. There was no one-to-one correspondence between aerial and terrestrial scene images. Further, each image was presented in an upright as well as inverted manner, resulting in a total of 720 images. In order to use all images of this database, the full set was divided into two groups, each containing 360 images, counterbalanced for categories and views. All images were presented centrally on the screen with a size of 768×768 pixels (subtending a visual angle of $18.2°$ horizontally and vertically).

2.4 Procedure

Subjects were informed that the purpose of the study was to investigate eye movement patterns in the perception of natural scenes and were asked to study the images in order to categorize each scene after the presentation. Subjects were randomly assigned to one of two image groups. To limit the total duration of the experiment, participants of both groups inspected only 240 scenes (40 from each category, 20 aerial and 20 terrestrial views, shown either upright ore inverted), selected from each group's image database (360 scenes). This procedure made sure that within each group the overlap of images between any two subjects was at least 66 % but there was no overlap between the groups. Each image was shown for 6500 ms, followed by a screen showing the six category labels arranged in a 2×3 grid. Category labels were randomly shuffled on a trial-by-trial basis, to avoid influences of responding to a preferred grid location (e.g., top left) across trials. Participants had to categorize the previous image by clicking on the appropriate category, and then the next trial started. The full experiment took 60 min in total to complete. An initial 9-point calibration and validation was performed before the start of each block, and calibration was checked prior to every 20th trial in the experiment.

2.5 Data Analysis

Raw eye movement data were preprocessed by removing fixations around eye blinks and outside the presentation screen, resulting in the inclusion of a total of 136, 406 fixations and saccades (86 % of the original data). Because of the positive skew in the distributions of fixation durations and saccade amplitudes, median rather than mean values were used as a measure of central tendency. For statistical testing the respective median values were subjected to repeated measures analyses of variance (ANOVA). All ANOVA factors were within-subjects, unless otherwise stated in the text. *Eta*-squared values are reported as estimates of the effect size (Levine and Hullett 2002).

3 Results and Discussion

3.1 Behavioral Measures

First we examined the percentage of correct responses for the four different types of scene presentations. The correct responses of all subjects were analyzed in a 2 (view: aerial, terrestrial) $\times$ 2 (rotation: upright, inverted) repeated measures ANOVA. We obtained a significant main effect for view, $F(1,28) = 43.43$, $p < 0.001$, $\eta^2 = 0.942$, in which viewers were more accurate for terrestrial than for aerial scene views (97.2 vs. 90.6), consistent with Loschky et al. (2010). However, there was no main effect for rotation, $F(1,28) = 1.35$, $p = 0.255$. This was due to the fact that there was a significant view $\times$ rotation interaction, $F(1,28) = 6.77$, $p = 0.015$, $\eta^2 = 0.052$, in which viewers showed slightly better classification performance for upright than inverted terrestrial images (97.7 vs. 96.7 %) whereas for aerial views the opposite was found (89.6 vs. 91.6 %). The lack of effect for image inversion on terrestrial scene categorization accuracy was likely due to the long stimulus presentation times of 6500 ms.

3.2 Eye Movement Measures

Our objective was to examine eye movement behavior during the free visual exploration of real-world scenes. Therefore, we systematically varied the scene viewpoint (aerial vs. terrestrial) and rotation (inverted vs. upright). In particular, we were interested in possible influences of these factors on fixation durations and saccade amplitudes at specific points in the time course of scene inspection. Furthermore, we investigated if scene viewpoint and rotation influence scene exploration behavior over time by comparing the eye movement scan paths as a function of these factors.

For the first analysis, fixation durations and saccade amplitudes were entered into two 2 (view: aerial vs. terrestrial) $\times$ 2 (rotation: inverted vs. upright) repeated measures ANOVAs, respectively. For fixation durations we observed significant differences for view, $F(1,28) = 5.33$, $p = 0.029$, $\eta^2 = 0.336$, and for rotation, $F(1,28) = 13.4$, $p = 0.001$, $\eta^2 = 0.4$. Fixation durations were longer for aerial views (329 vs. 320 ms) and longer if scenes were shown inverted (331 vs. 317 ms). There was also a significant interaction of view $\times$ rotation, $F(1,28) = 15$, $p < 0.001$, $\eta^2 = 0.264$, which was based on shorter fixation durations for terrestrial views if shown upright (309 vs. 326 ms), whereas no such difference was found for aerial views (both 331 ms). For saccade amplitudes we found significant differences for the influence of view, $F(1,28) = 22.21$, $p < 0.001$, $\eta^2 = 0.97$, but not for rotation, $F(1,28) = 1.14$, $p = 0.295$. Longer saccades were made during the exploration of aerial images (4.8 vs. 4.5 deg). No significant interaction was found, $F < 1$. Thus, the overall comparisons of fixation durations and saccade amplitudes as a function of view and scene orientation suggest that there are differences in gaze behavior depending on the type of scene people are looking at.

In the second analysis, we focused on particular periods of scene inspection to explore in detail the nature of the global differences reported above. First, we were interested in the earliest influences of the stimulus material, namely possible differences in gaze behavior immediately following the scene onset, which can be related to the processing of gist (i.e., the understanding of the overall meaning of a scene, see e.g., Oliva and Torralba 2006; Greene and Oliva 2009; Loschky and Larson 2010). Thus, we took a closer look at the fixations that started before the scene onset began, and ended after the scene onset. Those fixations were divided into the time before and after the scene onset. We predicted similar fixation durations for the time before the scene onset but expected differences for the fixation durations after the scene onset, if the content of a scene has an early influence on gaze behavior. The 2 (view: aerial vs. terrestrial) $\times$ 2 (rotation: inverted vs. upright) repeated measures ANOVA on the fixation times before the image onset revealed no differences, $F < 1$, while the same analysis on the remaining fixation time after the image onset demonstrated significant differences for view, $F(1,28) = 52.7$, $p < 0.001$, $\eta^2 = 0.9$, but not for rotation, $F(1,28) = 1.97$, $p = 0.171$. No significant interaction was found, $F(1,28) = 3.01$, $p = 0.093$. Thus, the remaining fixation time after the scene onset was larger for aerial than terrestrial scenes (239 vs. 214 ms). Finally, we analyzed the saccade amplitudes following the picture onset but found no differences with regard to the scene viewpoint, all $F < 1$.

In our final analysis, we were interested in the eye movement behavior (i.e., fixation durations and saccade amplitudes) over the time course of scene exploration. For this analysis, we divided the 6.5 s image presentation into three time periods: 0–2 s, >2–4.5 s, and >4.5–6.5 s. The first time period was from immediately following the scene onset to 2.0 s after onset. During this period, we assumed a more global scanning strategy (Antes 1974) which has been described in terms of ambient processing (Unema et al. 2005; Pannasch et al. 2008).

Consistent with previous studies, this interval included only eye movements up to 2 s after image onset (for this analysis we excluded the eye fixation spanning the scene onset and started with the first fixation after the image onset). The second time period was between these initial 2 s of scene exploration and up to 4.5 s, during which time we assumed there would be a transition from ambient to focal processing. The third time period was for the final 2 s of scene exploration, during which time we predicted a larger proportion of focal processing being more related to the extraction of details and particular object features.

Fixation Durations: We conducted a 3 (time period: first, second, third) $\times$ 2 (view: terrestrial, aerial) $\times$ 2 (rotation: upright, inverted) repeated measures ANOVA. For fixation durations we obtained significant main effects for time period, $F(2,56) = 46.9$, $p < 0.001$, $\eta^2 = 0.824$, view, $F(1,28) = 12$, $p = 0.001$, $\eta^2 = 0.082$, and rotation, $F(1,28) = 16.5$, $p = 0.005$, $\eta^2 = 0.045$. Furthermore, we observed a significant interaction for view $\times$ rotation, $F(2,56) = 12.1$, $p = 0.019$. No further interactions were found. Regarding the main effect for time period, post hoc testing revealed a significant increase in fixation durations across the time periods, from the first to the third (254 vs. 269 vs. 288 ms), all $p < 0.001$. Fixation durations were shorter for upright terrestrial scenes (260 ms) than for inverted terrestrial (272 ms) or either rotation of aerial views (both 274 ms; see Fig. 2, left panel).

Saccade Amplitudes: We conducted the same 3 (time period: first, second, third) $\times$ 2 (view: terrestrial, aerial) $\times$ 2 (rotation: upright, inverted) repeated measures ANOVA for saccade amplitudes and obtained significant main effects for time period, $F(2,56) = 69.4$, $p < 0.001$, $\eta^2 = 0.854$, and view, $F(1,28) = 13.5$, $p < 0.001$, $\eta^2 = 0.124$, but not for rotation, $F < 1$. No further interactions were found. The main effect for time period was due to the significant decrease of
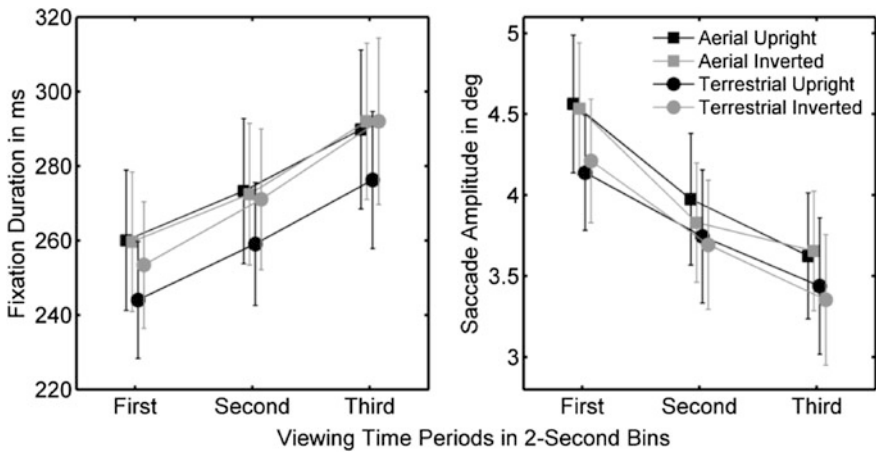
Fig. 2 Mean fixation durations (*left*) and mean saccade amplitudes (*right*) for view and rotation in the respective viewing time periods. The error bars represent 95 % confidence intervals

saccade length from the first to the last time period (4.36 vs. 3.81 vs. 3.52°), all $p < 0.001$. The second main effect was the result of longer saccades for aerial views (4.03 vs. 3.76°; see Fig. 2, right panel).

Finally, we examined the similarity of the fixation strategies for each view and rotation. Therefore, we used the ScanMatch algorithm (Cristino et al. 2010) which is based on the Needleman–Wunsch global sequence alignment algorithm (Needleman and Wunsch 1970) and creates letter sequences retaining fixation location, time and order information for each individual scene inspection. Pairs of sequences are then compared to find the optimal alignment between a pair. The similarity between two sequences is expressed by a normalized ScanMatch score, with similarity magnitude indicated by ScanMatch score distance from 0. The obtained similarity scores were averaged for each image and applied to a 2 (view: aerial vs. terrestrial) $\times$ 2 (rotation: inverted vs. upright) repeated measures ANOVA. A significant main effect was obtained for view, $F(1,179) = 26.1$, $p < 0.001$, $\eta^2 = 0.909$, but not for rotation, $F < 1$. No further interaction was found, $F(1,179) = 2.2$, $p = 0.14$. The main effect for view was due to viewers showing less similar fixation strategies for aerial views (0.548 vs. 0.569).

4 Conclusion

Our results revealed three major findings. First, fixation durations and saccadic amplitudes differed according to the respective view and orientation of the scenes. Fixations were longest for aerial views (both upright and inverted) and shortest for upright terrestrial views, suggesting greater difficulty in processing aerial than terrestrial scene images. Saccade amplitudes were larger for aerial than for terrestrial views, whether upright or inverted. Based on this observation, it would therefore seem that identifying aerial views requires the inspection of a larger proportion of image space. However, since we did not analyze the fixation locations in the image, the larger amplitudes could also result from repeated refixations of peripheral image regions. Thus, future work should incorporate the examination of fixation locations according to different fixation sequences within aerial versus terrestrial views (see, e.g., Çöltekin et al. 2010).

Second, the general gaze patterns over the time course of scene inspection followed earlier observations, namely there was a transition from early ambient to later focal processing, indicated by the increase in fixation durations and decrease in saccade amplitudes. However, fixation durations and saccade amplitudes are longer if the explored scene deviates from our everyday experience, namely terrestrial upright scenes. The longer fixation durations most likely indicate increased difficulty in processing and interpreting the visual information in the scene while the longer saccade amplitudes suggest intensified search for useful information in the scene. Future work should apply more advanced analytical methods, such as gaze map matching (Kiefer and Giannopoulos 2012) in order to provide further insight into these issues.

Finally, we analyzed additional parameters to obtain a better understanding of viewers' ongoing processing. For example, consistent with our predictions based on gist recognition for aerial scenes, we found longer first fixation durations (i.e., the remaining time of a fixation immediately after the picture onset) for both types of aerial views. This suggests that the difficulty in interpreting aerial scene images begins on the very first eye fixation on the image, which then influences further eye movements and visual processing while looking at the image. Furthermore, we determined the similarity of scan paths for each image across participants. Comparing the similarity indices for the different conditions revealed higher similarity between participants viewing terrestrial views compared to the exploration of aerial views. This may be due to viewers being better able to utilize normal viewing routines while looking at more familiar types of scenes, namely those taken from terrestrial views.

To summarize, our results reveal interesting differences in viewers' gaze patterns when inspecting the same information from our environment but presented from aerial versus terrestrial perspectives. In the case of aerial views, we found in various gaze parameters that this type of scene view has a clear influence on the balance of ambient and focal processing. While the general time course of ambient to focal processing remains stable, the proportion of focal processing is reduced and a greater dominance of the ambient mode is observed. The present study therefore provides insight into the general processing mechanisms involved in viewing complex imagery from different viewpoints. Future work is needed in order to assess how these basic mechanisms change as a function of expertise (e.g. Lloyd et al. 2002; Ooms et al. 2011).

References

Antes JR (1974) The time course of picture viewing. J Exp Psychol 103:62–70

Çöltekin A, Fabrikant SI, Lacayo M (2010) Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. Int J Geogr Inf Sci 24:1559–1575

Cristino F, Mathot S, Theeuwes J, Gilchrist ID (2010) ScanMatch: a novel method for comparing fixation sequences. Behav Res Methods 42:692–700

Findlay JM (1998) Active vision: visual activity in everyday life. Curr Biol 8:R640–R642

Greene MR, Oliva A (2009) The briefest of glances: the time course of natural scene understanding (Research Article). Psychol Sci 20:464–472

Kiefer P, Giannopoulos I (2012) Gaze map matching: mapping eye tracking data to geographic vector features. In: Proceedings of the 20th international conference on advances in geographic information systems, ACM, Redondo Beach, California, pp 359–368

Land MF (2004) The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. Exp Brain Res 159:151–160

Levine TR, Hullett C (2002) Eta-square, partial eta-square, and misreporting of effect size in communication research. Hum Commun Res 28:612–625

Lloyd R, Hodgson ME, Stokes A (2002) Visual categorization with aerial photographs. Ann Assoc Am Geogr 92:241–266

Loschky LC, Larson AM (2010) The natural/man-made distinction is made before basic-level distinctions in scene gist processing. Vis Cogn 18:513–536

Loschky LC, Ellis K, Sears T, Ringer R, Davis J (2010) Broadening the horizons of scene gist recognition: aerial and ground-based views. J Vis 10:1238

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Oliva A, Torralba A (2006) Building the gist of a scene: the role of global image features in recognition. Prog Brain Res 155:23–36

Ooms K, De Maeyer P, Fack V (2011) Can experts interpret a map's content more efficiently? In: Proceedings of the 25th international cartographic conference (ICC 2011), Paris, France, 3–8 July 2011

Pannasch S, Helmert JR, Roth K, Herbold A-K, Walter H (2008) Visual fixation durations and saccadic amplitudes: Shifting relationship in a variety of conditions. J Eye Mov Res 2(4):1–19

Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124:372–422

Unema PJA, Pannasch S, Joos M, Velichkovsky BM (2005) Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration. Vis Cogn 12:473–494

Velichkovsky BM, Joos M, Helmert JR, Pannasch S (2005) Two visual systems and their eye movements: evidence from static and dynamic scene perception. In: Bara BG, Barsalou L, Bucciarelli M (eds) Proceedings of the 27th conference of the cognitive science society. Lawrence Erlbaum, Mahwah, pp 2283–2288

Understanding Soil Acidification Process Using Animation and Text: An Empirical User Evaluation With Eye Tracking

P. Russo, C. Pettit, A. Coltekin, M. Imhof, M. Cox and C. Bayliss

Abstract This chapter presents a user study in which the participant performance is comparatively measured using two ways of presenting information: animation and text. The stimuli contain equivalent information, but use fundamentally different ways of communicating this information. We designed a workplace to simulate the process as it may occur in the real world. First, a representative task from an actual website was selected (i.e., understanding the soil acidification process). 50 participants first took part in a short 'study session', where they were told to remember as much as possible. Then they took a multiple choice test using either the animation or the text in an "open book" setting. The tested media have been assessed through the classical measures of *effectiveness* (error rate), and *efficiency* (time to complete the multiple choice test). Text users achieved a slightly higher score in the multiple choice test and required less time compared to animation users. In contrast, more of the animation users considered the questions "easy". Thus, against all intuition (yet in agreement with some of the previous findings in literature) animation does not appear to perform better for the tasks in this experiment. To further strengthen the experiment, an eye tracking study was also conducted with the animated displays for a more in-depth effort to explore user strategies when asked to 'remember as much as possible'.

P. Russo · C. Pettit
Faculty of Architecture, Building and Planning, The University of Melbourne, Parkville, VIC 3010, Australia

A. Coltekin (✉)
Department of Geography, University of Zurich, Zurich, Switzerland
e-mail: arzu@geo.uzh.ch

M. Imhof · M. Cox
Department of Primary Industries, Future Farming Systems Research Division, Parkville, VIC 3053, Australia

C. Bayliss
Melbourne eResearch Group, Faculty of Engineering, The University of Melbourne, Parkville, VIC 3010, Australia

1 Introduction

Animations are often used for visualising scientific concepts with spatial and temporal aspects in a dynamic form. Several studies have been conducted to compare animation versus other forms of representations including text (e.g., Tversky and Morrison 2002; Rebetez et al. 2010 and Lewalter 2003). For a variety of tasks, such as comprehension, learning, memory, communication and inference, previous research consistently shows that animations are not necessarily "better" than other forms of graphics or text (e.g., DiBiase et al. 1992; Tversky and Morrison 2002; Hegarty et al. 2003; Griffin et al. 2006; Harrower and Fabrikant 2008). Indeed, effective learning with animations depends on a number of factors, including the design and speed of animation and the prior knowledge of the users (Hegarty and Kriz 2008).

The design of the animation, for example, influences the internal cognitive processes and therefore how much users learn. Gog and Scheiter (2010) report that users learn better if spoken text (and not written) is integrated into animations. When written text accompanies an animation, users tend to read the text and largely ignore the pictorial information. Conversely, the presence of spoken text allows users to pay attention to the visual display. Seemingly, users' attention can also be better captured if the speed of the animation varies over time. For example, Mayer (2010) proposed that learners benefit more from an animation if its speed decreases with increasing playtime (and not vice versa). Furthermore, highlighting relevant features may also improve learning ability according to Mayer (2010).

The spatial ability and the prior knowledge of the learner may also influence learning outcomes when using animations as study material (Hegarty and Kriz 2008). Among others, Yang et al. (2003) demonstrate that people with high spatial abilities may have fewer problems in processing visualizations compared to people with lower abilities; thus potentially gaining a stronger benefit from using animations when learning. Similarly, recent eye tracking studies demonstrated that users with prior knowledge of the subject matter seem to distinguish more between relevant and irrelevant information by gazing faster and more often at the relevant display elements (Gog and Scheiter 2010; Çöltekin et al. 2010). Also, Mayer (2010) maintains that experts benefit more, in terms of learnability, from animations, compared to novices. Lowe (2003) suggests that the potential of animation as a learning tool may not be fully realised unless the design of such presentations

supports learners' extraction of domain-relevant information and its incorporation into existing knowledge structures.

Based on our (non-exhaustive) review of the literature, it appears likely that the contribution of animation in supporting learning outcomes depends also on the topic. In some studies, animations are reported to have slight advantages in comparison to static displays in communicating spatial aspects of dynamic processes since they support learners in constructing mental representations (Lewalter 2003). Various other studies also indicate that animation may be a useful tool for refreshing and retaining knowledge. Best results for comprehension have been found when using animation collaboratively (i.e., several viewers) and in combination with text (Rebetez et al. 2010). An overall superiority of animations over text on learning outcome is, however, clearly not evidenced.

We do not intend to discuss all such influencing factors in detail in this chapter. Rather, in this chapter, we demonstrate the challenges for an effective real world application of animation. In the process, we contribute to the body of knowledge in user studies with animation, by testing the efficiency of *information extraction* from an animation in comparison to text in a case study featuring a real world application. More specifically we pursue two research questions in this study: Firstly, "is it more efficient (and effective) to extract information from an animation or from a body of text describing the same animation in a complex real world learning task?" (RQ1) and secondly, "do the eye movement measurements recorded during the inspection of the animation underpin in any way the performance measurements of RQ1?"(RQ2).

The real world application utilised in this study has been designed to enhance understanding of the soil acidification process. Soil acidification can negatively impact a number of natural resources and the process itself is a complex environmental process which requires understanding many chemical concepts as well as interactions within and between them. The featured animation is created by soil science experts from the Department of Primary Industries and made available through the Victorian Resources Online (VRO) website for public use: http://www.dpi.vic.gov.au/vro. Visualisations such as these provide a means for capturing expert knowledge in an explicit way to be used for knowledge transfer (Imhof et al. 2010) and science communication. The text stimulus is provided on the same website; and this is because, according to the Whole of Victorian Government Website Standard; "to assist people that may have impeded access to the web, graphical elements, such as animations need to be accompanied with an accessible alternative (e.g. Word, text, RTF or HTML) that appropriately describes the content shown".[1]

---

[1] http://www.egov.vic.gov.au/victorian-government-resources/website-management-framework-wmf-/government-website-standards-victoria/whole-of-victorian-government-website-standards-overview.html

2 Experiment

2.1 Design

To balance experimental control with ecological validity, we simulated a work-place in which people use the provided stimuli (i.e., animation and text) to eventually answer questions on the soil acidification process. The experiment consisted of three sections (Fig. 1). The Sect. 1 was a 4-min training session. Half of the participants studied the animation[2] and the other half the text (Appendix 1). The content of the stimuli was explained to the participants before the beginning of the training session and the participants were asked to memorise as much information as possible, including the location of information within the stimuli. In Sect. 2 participants completed a multiple choice (MC) test answering questions relevant to the content of the stimuli with a time limit of 10-min. For this task, participants were allowed to use the stimulus as in an 'open book exam', i.e., the users were allowed to look at the stimulus as often as they needed. This behaviour is similar to how the online user population is expected to utilise such stimuli. In Sect. 3, participants' background and satisfaction with the stimulus were assessed by a brief questionnaire.

2.2 Participants

The group of participants comprised 50 students (27 male, 23 female, average age of 22) from the University of Melbourne (under- and postgraduates, PhD students). None reported any colour vision problems, which would have excluded them from the test as this may have caused a bias. 40 % of the participants reported beginner's knowledge of soil chemical processes and 10 % indicated that they had studied or worked in related fields (Table 1).

2.3 Stimuli

The stimuli contain the exact same information though their representation differs fundamentally. Comparing *informationally equivalent* media with differences in representation is a fairly common practice (Simon and Larkin 1987; Coltekin et al. 2009).

   The animation depicts soil processes in the context of a dairying agro-eco-system. It has been developed to convey complex processes to a broad target

---

[2] http://vro.dpi.vic.gov.au/dpi/vro/vrosite.nsf/pages/soilhealth_acidification
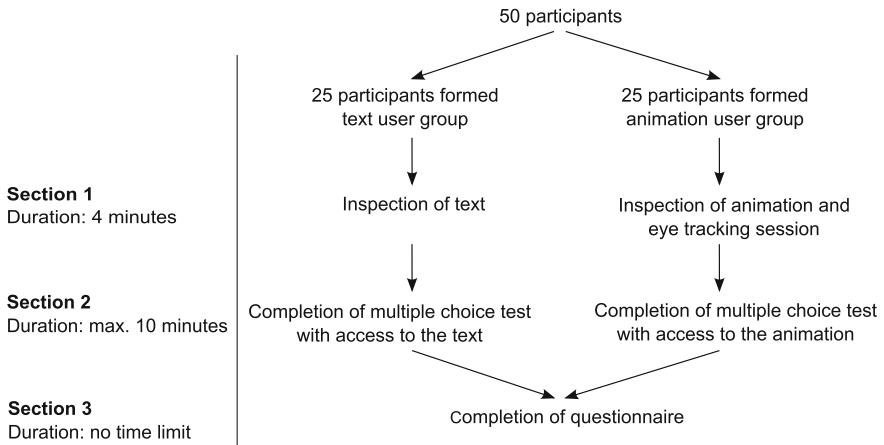
Fig. 1 Design of the experiment

Table 1 Summary of the data collected during the experiment

	Animation	Text
Number of participants	25	25
Mean participants' age	24.1	20.4
Multiple choice test		
Mean number of correct answers	8.52	8.60
Required time (min)	06:17	05:38
Questionnaire		
Prior knowledge about chemical processes in soils (1 = no knowledge, 6 = high knowledge)	2.6	2.8
Participants rating information extraction 'very easy'	4	2
Participants rating information extraction 'easy'	17	15
Participants rating information extraction 'difficult'	3	5
Participants rating information extraction 'very difficult'	0	1

audience e.g., researchers, administrators, educators, students and the general public. The animation was created from 'storyboards'—a series of hand-drawn sketches outlining all the successive events in the animation—derived from workshop sessions with subject matter specialists. An entire scene was initially created that included all the graphical elements from each of the single story-boards. This was then Web-enabled using Adobe Flash to create a sequence of animated scenes that users can move through scene-by-scene or play in a con-tinuous animation (with pre-set timing intervals). The animation contains various graphical elements (10 scenes), text and voice-over. The information provided by the written and spoken text was identical. The ability to switch the audio off was disabled during the experiment, thus the animation was always coupled with sound.

The other stimulus (text) contained the same information as the animation and based on the animation transcript.[3] In the experiment it has been provided as an MS Word document and included only written text of almost one page length without any figures (Appendix 1).

2.4 Tasks

The MC test was based on 10 questions on factual and procedural knowledge presented (Appendix 2) in the animation and text (Lewalter 2003). For each question, three answers (two incorrect and one correct) were provided. Participants were asked to select the correct answer. For the completion of the MC test, participants had a maximum of 10 min time and were allowed to use the stimulus inspected in Sect. 1. Thus, the upper half of the screen used by the participants was covered by the multiple choice test, whilst either the animation (Fig. 2a) or the text (Fig. 2b) was displayed on the lower part of the screen.

After the MC test, a questionnaire was provided to the participants. Participants were asked whether they needed to look at the stimulus to complete the MC test in Sect. 2 or whether they remembered the information. Participants were also requested to rate the degree of ease or difficulty in undertaking the tasks using a 4-point Likert scale (i.e. 'very easy', 'easy', 'difficult', 'very difficult') (Komorita 1963). The same questionnaire was used to collect information on the participants, including their age, gender, field of study and whether they have any visual impairment. They also assessed their prior knowledge of soil chemical processes using a 6-point Likert scale ranging from 1 (=no knowledge) to 6 (=high knowledge) (Appendix 3).

2.5 Procedure

After welcoming the participants, the experimenter provided a brief introduction to the experiment and the eye tracking system. All participants signed a consent form for video and eye movement recording (text users were also recorded but not analysed in the scope of this chapter). Before recording began, participants were instructed to assume a comfortable position and not move too much during the experiment, to maximize the eye movement recording accuracy. A calibration with the eye tracker followed. Participants were then provided with a mouse to scroll up and down the text and a keyboard to mark the answers in Sect. 2. During the entire test, participants wore headphones to reduce background noise and to assist listening to the spoken text in the animation (Fig. 3).

---

[3] http://vro.dpi.vic.gov.au/dpi/vro/vrosite.nsf/pages/soilhealth_acidification_transcript
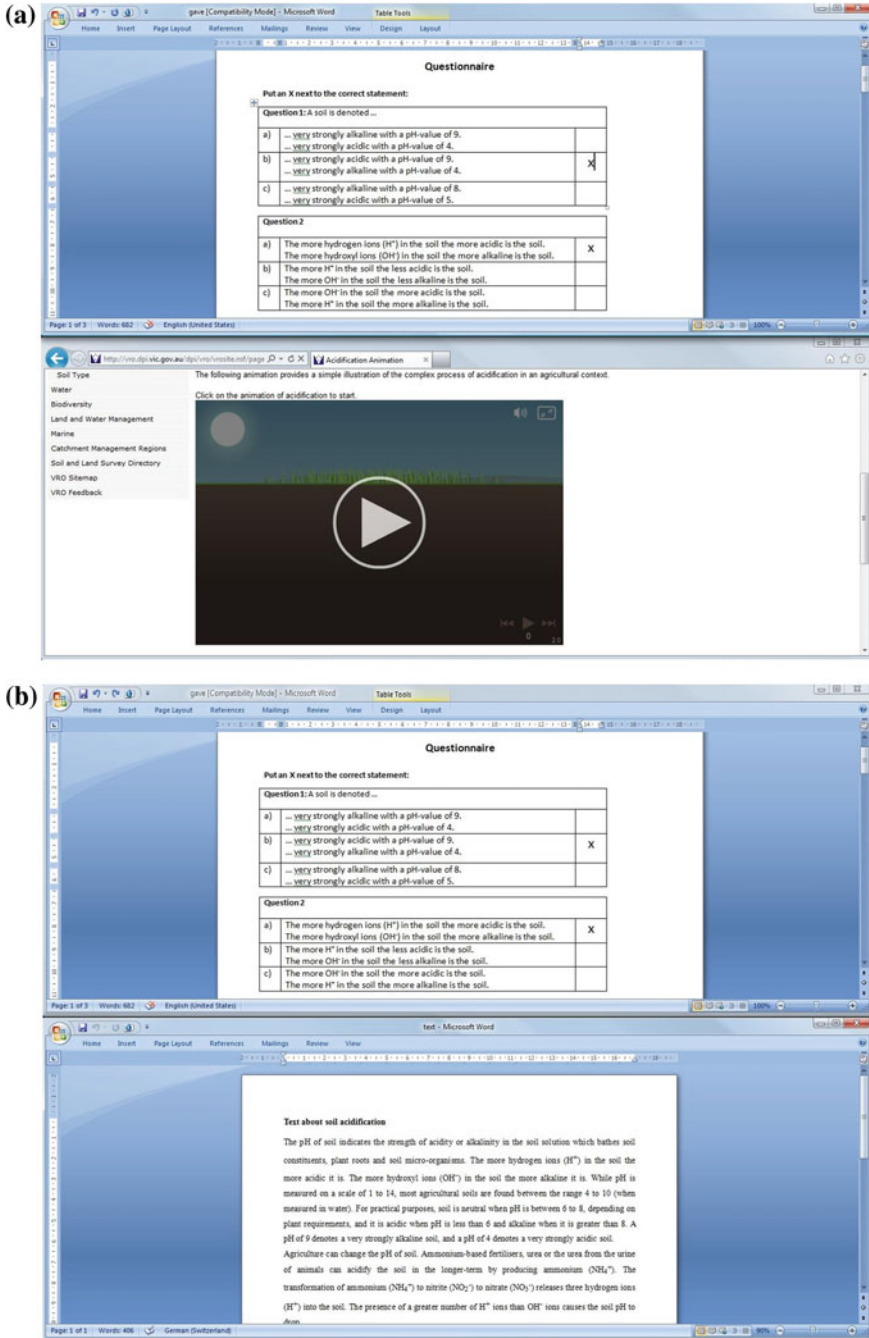
Fig. 2 **a** Display of multiple choice test with animation **b** Display of multiple choice test with text

Fig. 3 Experimental hardware set-up

The experimenter ensured that participants inspected the stimulus for 4 min in the first section and that the second section did not exceed 10 min. As an incentive to undertake the experiment a 20 AUD voucher was given to each participant.

3 Methodology

Usability metrics. The methodology is strongly linked to our two research questions, i.e., "is it more efficient (and effective) to extract information from an animation or from a body of text describing the same animation in a complex real world learning task?" (RQ1) and "do the eye movement measurements recorded during the inspection of the animation underpin in any way the performance measurements of RQ1?" (RQ2). For the first question, we can resort to most classical usability evaluation metrics of satisfaction, effectiveness (accuracy) and efficiency (speed) (Nielsen 1993). In fact, efficiency as we intend in RQ1 is closely linked to effectiveness and can be defined as the relation between the accuracy with which a task is completed and the time required to complete that task

(Frokjaer et al. 2000). Accordingly, the performance measurements to assess RQ1 are (i) the number of correct relative to incorrect answers chosen in the multiple choice test, and (ii) the required time to complete the multiple choice test. Questionnaire responses on the difficulty of extracting information from the stimulus should reveal the level of difficulty experienced during the tasks.

Eye tracking. The eye movement measurements report the location and duration of the gaze. In this study we consider a common eye movement metric, the *Mean Fixation Duration* (MFD), on specifically defined Areas Of Interest (AOI) within each scene of the animation (Schmidt-Weigand et al. 2010) (Fig. 4). According to Jacob and Karn (2003) a fixation (where the gaze is relatively still) is typically defined by a dispersion threshold of $\sim 2°$ and minimum duration of ~ 100–200 ms, however, it is necessary to note that fixation threshold is currently arbitrarily defined in literature and varies from 50 to 500 ms (Coltekin et al. 2009). In this study, we adopt a minimum duration of 144 ms for defining a fixation. This value is within the window of fixation thresholds that are commonly reported in usability studies with eye tracking. Figure 5 shows a sequence of fixations. The examination and interpretation of the measurements is based on Jacob and Karn (2003) and assumes that the longer the MFD, the more effort participants have exerted to extract information from the AOI. In relation to RQ2; we analysed if there is any correlation between the time the participants took to complete the multiple choice test and the MFD on the AOIs that contain the information to answer the questions correctly ("relevant AOIs"). That is, if animation users
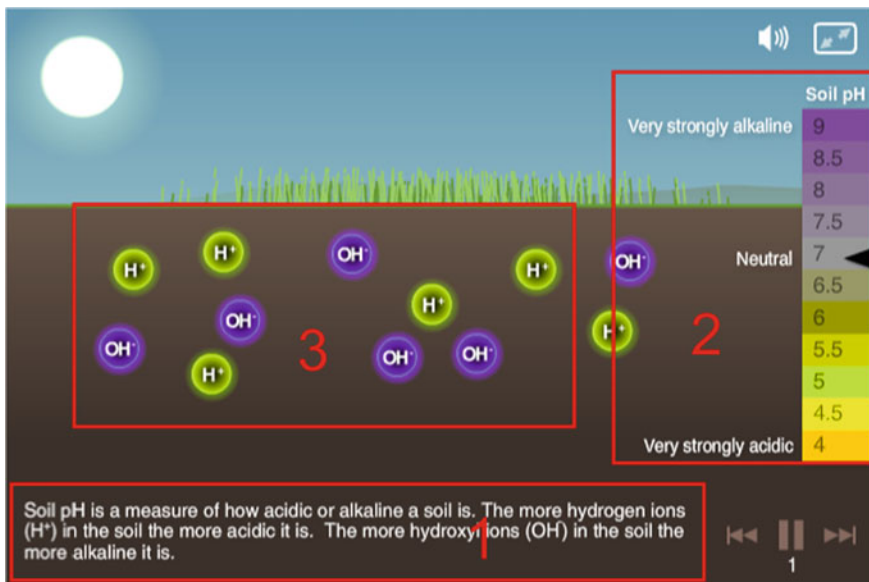
Fig. 4 Scene 1 of the animation and the chosen AOIs (outlined as *rectangles* and numbered as *1*, *2* and *3*)
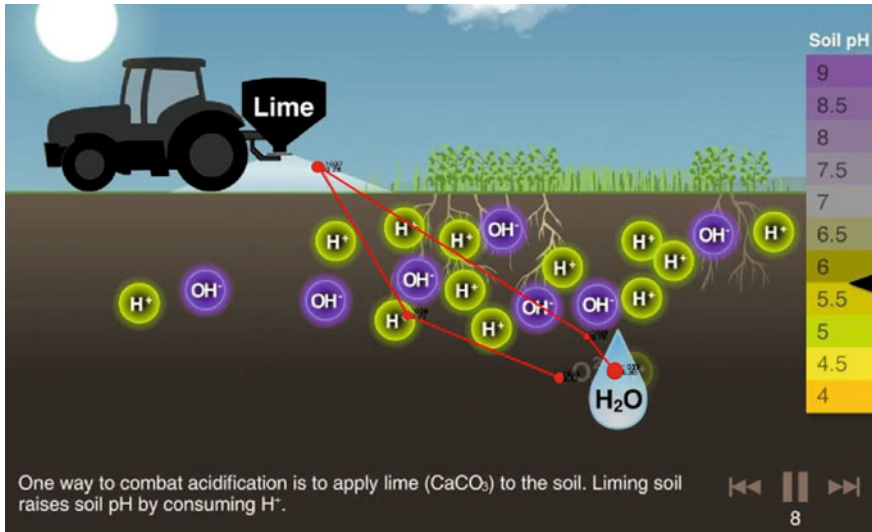
Fig. 5 An example (the trajectory starts at the water drop)

required more time to complete the MC test, we studied the eye movements in the training session that took place before the MC test. If the MFD on the relevant AOIs were relatively high, then we interpret that animation users may have been delayed by difficulties encountered in decoding the information.

4 Results

4.1 Multiple Choice Test and the Questionnaire

Table 1 shows the results of the MC test and questionnaire. On average, the text users have achieved a slightly higher score of correct answers compared to the animation users. The mean time to complete the MC test was about 40 s less for text users. In contrast, more of the animation users stated the extraction of information 'very easy' and 'easy'.

While the text users appear to have a slightly higher prior knowledge on the topic, nearly all participants (47 out of 50) have returned to the stimulus as they were responding to the questions. Thus, at this point, we do not further analyze the differences among participants in relation to their prior knowledge.

To better understand the differences between the two stimuli, a sum of rank test (Mann–Whitney U test) was carried out to statistically test the distribution of correct answers. The required time has been analysed with a t test for equality of means. For both methods, a decision criterion of $p = 0.05$ has been applied. The

Mann–Whitney U test revealed ($p = 0.895$) that the difference in performance between animation and text users is not statistically significant. Because time data are typically non-normally distributed, the required times were first logarithmically transformed before statistically analysed (Pettersson et al. 2009). According to the t-test there was no significant difference between the two stimuli for the completion time of the MC ($p = 0.228$).

4.2 Eye Tracking

According to the initial tests, we observe a consistent difference in favour of text in our findings in terms of efficiency (Table 1). However, since the results are not statistically significant, the findings are inconclusive. To strengthen the experiment, we recorded eye movements with animation users in Sect. 1. In our limited eye movement analysis, we investigate if the animation users have spent more time (fixated longer) or as much time (fixated equally) in 'irrelevant' parts of the visualizations as they did in the 'relevant' parts during the training session. Note that the participants did not know what was relevant or irrelevant for the tasks at the time of recording, but were asked to 'remember as much as possible' (similar to studying for an exam). We defined *relevant AOIs* as those that contain necessary information for answering the MC questions. *Irrelevant AOIs* are selected areas that contain information which is not necessary for answering the MC questions. Once AOIs were defined and classified, we compared the MFDs between 'relevant' and 'non-relevant' AOIs. Table 2 shows the MFDs for each scene in relation to the relevant/non-relevant AOI in seconds. The numbers in the column *Question* show which of the 10 questions of the MC test is answered with this scene and AOI. For example, AOI 2 of Scene 1 (Fig. 4) answers question 1 of the MC test (Appendix 2).

The results (Table 2) demonstrate the MFDs for relevant AOIs are not higher than the MFDs for the non-relevant AOIs in any case. The mean MFD of the relevant AOIs (0.36 s) is lower than the mean MFD of the non-relevant AOIs (0.40 s) (Table 3). In other words, the relevant AOIs have been fixated for a shorter time than the non-relevant AOIs in average. The mean MFDs have been statistically analysed with a t-test for equality of means with a decision criterion of $p = 0.05$. The test revealed ($p = 0.160$) that there is no significant difference between the mean MFDs of the two AOIs. Based on these findings, one possible interpretation is that the animation users have not been delayed due to difficulties in extracting the required information, but rather may have been distracted by (or attracted to) the other elements that are displayed on the studied scenes.

Table 2 Eye tracking measurements (MFD) for each scene and relevant/non-relevant AOI

Scene of animation	AOI		MFD (s)	Question
	Non-relevant	Relevant		
Scene 1			0.37	
		1	0.36	2
		2	0.35	1
	3		0.39	
Scene 2			0.36	
		1	0.33	3,6
	2		0.50	
	3		0.36	
Scene 3			0.37	
		1	0.30	7
	2		0.44	
		3	0.38	7
Scene 4			0.38	
	1		0.33	
	2		0.37	
		3	0.39	8
Scene 5			0.41	
		1	0.35	4
	2		0.73	
		3	0.40	4
Scene 6			0.36	
		1	0.33	5
	2		0.30	
		3	0.37	5
Scene 7			0.39	
	1		0.32	
	2		0.37	
	3		0.45	
Scene 8			0.39	
	1		0.35	
	2		0.40	
		3	0.41	10
Scene 9			0.40	
		1	0.30	9
	2		0.43	
	3		0.45	
Scene 10			0.46	
	1		0.33	
	2		0.24	
	3		0.47	

Table 3 Mean MFD of relevant/non-relevant AOIs

Mean MFD (s) of relevant AOIs	Mean MFD (s) of non-relevant AOIs
0.36	0.4

5 Conclusions and Outlook

The differences between the animation and text users' performance in this study is not statistically significant, thus essentially the experiment remains inconclusive in terms of a 'clear winner'. Nonetheless, we observe a trend in the results which slightly favours text over animation, both in the number of accurate responses and in the time required to complete the tasks. Despite the 'feeling' that animations seem to induce in the viewers that information is easier to grasp (as a trend, animation users have rated the tasks 'easier' compared to those who used text), animation users did not seem to perform better within the limits of this study. The perceived difficulty with text, or rather the perceived ease with animation, remains unexplained. This may be because of the tedious appearance of the text and the visual appeal of the animation. Qualitative feedback by the participants supports this; the text has been described as 'dry' and 'boring' by some. Upon closer examination, the animation is visually highly demanding due to frequently changing details and multiple AOIs which may cause distraction and lead to inefficiency while trying to extract information. Also, the animation contains more interruptions (i.e. scene changing) while the text is more fluent which may facilitate comprehension. Furthermore, the information in the animation is more fragmented among the three sources (spoken and written text and image). In this regard, a participant stated that the three sources were 'competing'. We surmise that this may be due to the availability of three integrated sources of information making it difficult to focus on any one of them. The text may be more monotonous for many users but the information is extracted from it seemingly 'just as well', perhaps due to the simple structure of the text and with fewer opportunities for users to explore the information in a non-linear way (which is possible with the animation). These are possible explanations for the greater time required of animation users which suggests that further in-depth investigations will be useful.

This study, however, has shown controversy results within the Satisfaction, Effectiveness, Efficiency (SEE) analysis with satisfaction not correlating with effectiveness and efficiency. Users seem to be more stimulated to use the animation because of its visual appeal, but the extraction of information from text seems to be no less efficient if we look at the trends (also see e.g., Tversky et al. 2002). Participants' misguided confidence based on the qualities of the stimulus has been also reported in earlier studies (e.g., Zanola et al. 2009). A more in-depth inspection to study this conflict is worthwhile as other findings in literature suggest animation to be a complex form of visualisation that does not have facilitating effects as expected. Explanations for animation users' delay in completing the tasks can also not be intuited from the eye tracking results. The descriptive results indicate that animation users spend more time in 'irrelevant' parts of the display. These results encourage further research on the potential contribution of eye movement measurements for explaining efficiency differences between media.

At this stage, we have early indications from the observations in this experiment that allow us to build hypotheses. In follow-up research, modifications may be

undertaken on the animation to further strengthen the current study and understand better if/when animations are good 'learning material' in real world applications for science communication and learning. As earlier literature indicates, the efficiency of animated displays is very likely task dependent, thus further tests with a variety of (ideally taxonomized) learning tasks should be designed. Display complexity also plays a clear role and there are numerous possibilities to vary in the visualisation design. For example, incorporation of buttons for switching written text and audio on/off, changing animation speed, highlighting thematically relevant features (Mayer 2010; Harrower and Fabrikant 2008), and simplifying animations to contain only one specific AOI would be all valuable to test the various aspects that are relevant to further understanding the potential of animated displays as 'learning material'. Thus, it would be useful to analyse if such changes increase the performance of the animation users in respect to the time required for completing the multiple choice test. Additional testing could also be undertaken on the value of each type of media for facilitating longer term knowledge retention by users and improving their understanding of complex soil processes.

Appendices

A.1 Appendix 1: Text Used in the Experiment

Soil Acidification

The pH of soil indicates the strength of acidity or alkalinity in the soil solution which bathes soil constituents, plant roots and soil micro-organisms. The more hydrogen ions (H^+) in the soil the more acidic it is. The more hydroxyl ions (OH^-) in the soil the more alkaline it is. While pH is measured on a scale of 1–14, most agricultural soils are found between the range 4 and 10 (when measured in water). For practical purposes, soil is neutral when pH is between 6 and 8, depending on plant requirements, and it is acidic when pH is less than 6 and alkaline when it is greater than 8. A pH of 9 denotes a very strongly alkaline soil, and a pH of 4 denotes a very strongly acidic soil. Agriculture can change the pH of soil. Ammonium-based fertilisers, urea or the urea from the urine of animals can acidify the soil in the longer-term by producing ammonium (NH_4^+). The transformation of ammonium (NH_4^+) to nitrite (NO_2^-) to nitrate (NO_3^-) releases three hydrogen ions (H^+) into the soil. The presence of a greater number of H^+ ions than OH^- ions causes the soil pH to drop.

Nitrate leaching, where NO_3^- moves below the root zone and cannot be used by plants is a significant source of agricultural acidification. Furthermore if the ammonium (NH_4^+) living in nodules on legume roots is not all used up by the crop or pasture the soil can become more acidic. Agriculture can also accelerate acidification by removing alkaline products such as wool, milk, cereal grain, legumes and hay. The reverse is also true, where the introduction of manure,

decaying animals, silage and stockfeeds can add alkalinity back into the soil and therefore increase soil pH.

Plant production can be constraint on strongly acid soils by aluminium toxicity and manganese toxicity. Both are more soluble at low pH, for example, aluminium dissolves into the soil solution as Al^{3+} that is taken up by the plant causing root deformation and stunted plant growth.

If a soil continues to acidify until it becomes very strongly acidic, biological activity, soil structure and nutrient toxicity and deficiency can become significant challenges to productive agriculture. One way to combat acidification is to apply lime ($CaCO_3$) to the soil. The breakdown of lime ($CaCO_3$) in the soil produces oxygen (O_2^-)·and water (H_2O). This reaction consumes H^+ ions and increases soil pH.

A.2   Appendix 2: Multiple Choice Test

Put an X next to the correct statement:

Question 1: A soil is denoted…
(a)   ... Very strongly alkaline with a pH-value of 9.
                              ... Very strongly acidic with a pH-value of 4.
(b)   ... Very strongly acidic with a pH-value of 9.
                              ... Very strongly alkaline with a pH-value of 4.
(c)   ... Very strongly alkaline with a pH-value of 8.
                              ... Very strongly acidic with a pH-value of 5.

Question 2
(a)   The more hydrogen ions (H^+) in the soil the more acidic is the soil.
           The more hydroxyl ions (OH^-) in the soil the more alkaline is the soil.
(b)   The more H^+ in the soil the less acidic is the soil.
           The more OH^- in the soil the less alkaline is the soil.
(c)   The more OH^-in the soil the more acidic is the soil.
           The more H^+ in the soil the more alkaline is the soil.

Question 3
(a)   The use of fertilisers and urea increase soil pH.
(b)   The use of fertilisers and urea decrease soil pH.
(c)   The use of fertilisers and urea decrease soil pH if air temperature is higher than 20 °C.

Question 4
(a)   By harvesting hay, grain and legumes alkaline products are removed which decreases the pH-value.

(continued)
(b) By harvesting hay, grain and legumes alkaline products are removed which increases the pH-value.
(c) By harvesting hay, grain and legumes acidic products are removed which increases the pH-value.

Question 5
(a) With the introduction of manure, decaying animals, silage and stockfeeds acidic products are added to soils which increases the pH- value.
(b) With the introduction of manure, decaying animals, silage and stockfeeds alkaline products are added to soils which increases the pH-value.
(c) With the introduction of manure, decaying animals, silage and stockfeeds alkaline products are added to soils which decreases the pH- value.

Question 6
(a) Agriculture has only acidifying effects on the pH of soil.
(b) Agriculture is an important factor of influence on pH of soil since it can accelerate acidification but also add alkalinity.
(c) The acidifying and alkaline effects of agriculture on soil compensate in order to not change the pH of soil.

Question 7: The transformation of ammonium (NH_4^+) to nitrite (NO_2^-) to nitrate (NO_3^-)…
(a)   ... releases OH^- ions raising soil pH.
(b)   ... releases H^+ ions raising soil pH.
(c)   ... releases H^+ ions reducing soil pH.

Question 8: NH_4^+ on Iegurne roots and NO_3^- leached below root zone because it is not used by animals and plants
(a)   ... can lead to increased soil alkalinity.
(b)   ... causes soil acidification.
(c)   ... does not change soil pH.

Question 9: Aluminium toxicity dissolves…
(a) ... easier.a.t low pH, The product Al3+ is taken up by the plant constraining it production.
(b) ... easier at high pH, The product Al3+ is taken up by the plant constraining it production.
(c) ... easier at low pH, The product Al3+ defenses plant against lice attacks.

Question 10: The dissolution of lime ($CaCO_3$)
(a)   ... comsumes H^+ ions resulting in the soil become more alkaline.
(b)   ... comsumes H^+ ions resulting in the soil become more acidic.
(c)   ... comsumes OH^- ions resulting in the soil become more alkaline.

A.3 Appendix 3: Questionnaire

1. Did you use the animation/text to respond to the questions? (Yes/No) lf yes:
2. How easy was it to find the answers in the animation or in the text (very easy/easy/difficult/very difficult)?

   Participant's background questions:

1. Age:
2. Gender (m/f):
3. Do you have any visual impairment? (If yes, please describe the constraints)
4. How much do you know on a scale from 1 (no knowledge) to 6 (high knowledge) about chemical processes in soils?
5. Do you study or work in the field(s) of:

   - Agriculture
   - Soil science
   - Chemistry
   - Biology
   - others

References

Çöltekin A, Heil B, Garlandini S, Fabrikant SI (2009) Evaluating the effectiveness of interactive map interface designs: a case study integrating usability metrics with eye-movement analysis. Cartogr Geogr Inf Sci 36(1):5–17

Çöltekin A, Fabrikant SI, Lacayo M (2010) Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. Int J Geogr Inf Sci 24(10):1559–1575

DiBiase D, MacEachren AM, Krygier JB, Reeves C (1992) Animation and the role of map design in scientific visualization. Cartogr Geogr Inf Sci 19(4):201–214

Frokjaer E, Hertzum M, Hornbaek K (2000) Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 345-352

Gog T, Scheiter K (2010) Eye tracking as a tool to study and enhance multimedia learning. Learn Instr 20(2):95–99

Griffin AL, MacEachren AM, Hardisty F, Steiner E, Li B (2006) A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. Ann Assoc Am Geogr 96(4):740–753

Harrower M, Fabrikant SI (2008) The role of map animation for geographic visualization. In: Dodge M, McDerby M, Turner M (eds) Geographic visualization. Wiley, Chichester, pp 573–605

Hegarty M, Kriz S (2008) Effects of knowledge and spatial ability on learning from animation. In: Lowe R, Schnotz W (eds) Learning with animation: research implications for design. Cambridge University Press, New York, pp 3–29

Hegarty M, Kriz S, Cate C (2003) The roles of mental animations and external animations in understanding mechanical systems. Cogn Instr 21(4):325–360

Imhof M, Cox M, Fadersen A, Harvey W, Thompson S, Rees D, Pettit C (2010) Natural resource knowledge and information management via the Victorian resources online website. Future Internet 3(4):248–280

Jacob RJK, Karn KS (2003) Eye tracking in human–computer interaction and usability research: ready to deliver the promises. In: Hyona J, Radach R, Deubel H (eds) The mind's eye: cognitive and applied aspects of eye movement research. Elsevier Science, Amsterdam, pp 573–605

Komorita SS (1963) Attitude content, intensity, and the neutral point on a Likert scale. J Soc Psychol 61(2):327–334

Larkin JH, Simon HA (1987) Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 11(1):65–100

Lewalter D (2003) Cognitive strategies for learning from static and dynamic visuals. Learn Instr 13(2):177–189

Lowe RK (2003) Animations and learning: selective processing of information in dynamic graphics. Learn Instr 13:157–176

Mayer RE (2010) Unique contributions of eye-tracking research to the study of learning with graphics. Learn Instr 20(2):167–171

Nielsen J (1993) Usability engineering. Morgan Kaufmann, San Francisco

Pettersson LW, Kjellin A, Lind M, Seipel S (2009) On the role of visual references in collaborative visualization. Inf Visual 9(2):98–114. doi:10.1057/ivs.2009.2

Rebetez C, Bétrancourt M, Sangin M, Dillenbourg P (2010) Learning from animation enabled by collaboration. Instr Sci 38(5):471–485

Schmidt-Weigand F, Kohnert A, Glowalla U (2010) A closer look at split visual attention in system- and self-paced instruction in multimedia learning. Learn Instr 20:100–110

Tversky B, Bauer J, Betrancourt M (2002) Animation: can it facilitate? Int J Hum Comput Stud 57(4):247–262

Yang E, Andre T, Greenbowe TJ, Tibell L (2003) Spatial ability and the impact of visualization/ animation on learning electrochemistry. Int J Sci Edu 25(3):329–349

Zanola S, Fabrikant SI, Çöltekin A (2009) The effect of realism on the confidence in spatial data quality in stereoscopic 3D displays. In: Proceedings of the 24th international cartography conference (ICC 2009), Santiago, Chile

Part IX
Cartography and GIS in Education

Perspectives on Developing Critical Human GI Capacity in a Developing Country Context

Felicia O. Akinyemi

Abstract The geospatial industry is experiencing unprecedented growth despite the global financial maelstrom. Consequently, the demand for manpower skilled in geographic information science (GISc) and technologies is great and this holds true both in developed and developing countries. Projections globally is that of a shortfall in the supply of geospatially skilled manpower. There is therefore the urgent need to develop a critical mass of geospatial specialists. In a developing country context as Rwanda, the awareness is high of the use of geographic information (GI) for economic development. Currently, every ministry is seeking to produce location based data and information in their respective domains. This is exemplified by the fact that GIS positions were recently created in the ministries. With increasing demand for GI personnel in both public and private sectors, there is need to develop human GI capacity. This chapter describes strategies currently being used in geospatial manpower development in Rwanda. It reviews the state of available programs vis-a-vis the areas of need with the aim of highlighting gaps in existing curriculum. In addition to the regular or traditional model of course delivery, other options for training and educating geospatial personnel are discussed and recommendations made based on lessons learnt.

F. O. Akinyemi (✉)
Faculty of Architecture and Environmental Design, Kigali Institute of Science and Technology, Kigali, Rwanda
e-mail: felicia.akinyemi@gmail.com

1 Introduction

Despite the global economic maelstrom of the last couple of years, the geospatial industry is fast growing. Worldwide, the industry is forecast to grow at a Compound Annual Growth Rate (CAGR) of 9.2 percent over the period 2011–2015. The GIS market in Retail Sector is forecast to reach $456.5 million in 2014. One of the key factors contributing to this market growth in retailing is the increasing adoption of GIS to site stores and business outlets. The general rise in demand for geospatial products and services is driven by an increasing global need for location based data and information in both developed and developing countries (Tech-Navio 2011, 2012). Ranging from countries launching their own earth observation and/or global positioning systems to others modernizing their mapping strategies, the development and use of GI abound. GIS data is today the fastest growing segment of the geospatial business. GIS Data has grown at a compound annual rate of 15.5 % for the last 8 years, about twice the rate of growth for software and services (Foundyller 2011). This is due to increasing demand for GIS from governments, businesses and militaries (TechNavio 2012). The global geospatial industry is now a very important contributor to the world economy, particularly in the engineering application sector.

Developments in satellite based mapping, the availability of lots of data as well as the widespread use of global positioning system (GPS) devices, coupled with the availability of geo-browsers such as Google Earth (http://www.google.com/earth/index.html), Google Maps (https://maps.google.com), ESRI ArcGIS Explorer (http://www.esri.com/software/arcgis/explorer/download), NASA World Wind (http://worldwind.arc.nasa.gov/java/), and Microsoft Bing Maps (http://www.bing.com/maps/) have popularised geospatial products.

The GI user base is ever-growing and consequently, we are experiencing increasing adoption of GIS technologies in different fields. This widespread adoption of geospatial technologies also increases the demand for geospatially skilled personnel (Ramli et al. 2010). Though the context might differ, the consensus is that there is a lack of skilled geospatial manpower in several countries. Some examples are Ethiopia (Gondo and Zibabgwe 2010), Uganda (Republic of Uganda 2010), India (Government of India 2010), Sweden (www.sweden.se/eng/Home/Work/Get-a-job/Labor-shortage-list/), US (DiBiase 2012).

The foregoing reveals that human capacity building is critical for success in utilising geographic information and communications technologies (Geo-ICT). Invariably, there is the need to use the latest GIS, remote sensing and database technologies to support research, decision making and development (Schilling et al. 2005). To man these Geo-ICT applications and meet the increasing demand for GI in Rwanda, education and training are essential to produce a critical mass of geospatially skilled manpower. This chapter examines strategies used in developing GI personnel locally. It reviews available programs, industry GI demand and areas where existing curricula must focus. Options for training and educating geospatial personnel are discussed and lastly, some recommendations are made.

2 Locally Developing Needed Geospatial Manpower

Rwanda seeks to increase its citizen's quality of life by achieving rapid and sustainable economic growth with ICTs playing a major role (MINECOFIN 2012). As the spatial dimension to addressing complex national development issues is known, most Rwandan institutions aspire to use Geo-ICTs (GIS, remote sensing, cartography, database management technologies), in their day-to-day activities (Akinyemi and Uwayezu 2011).

GI applications in Rwanda is increasing partly due to the government's use of evidence-based decision-making which feeds scientifically sound results into policy (Akinyemi 2012). Consequently, the demand is high for geospatial personnel. Majority of these personnel are being trained abroad with a full-time bachelors program requiring 4 years, masters requires 18 months, a doctorate requires 3 years minimum. This is a long waiting time before the employee can return to work. This predisposes that educating and training geospatial personnel locally is most desirable. Arguably, needed manpower can be best developed locally in order to shorten the time employers have to wait before employees are back to work from studies. Also, available geospatial personnel should be effectively engaged, for example, they can share their time across organizations.

3 Strategies to Building Needed GI Capacity

The near absence of GI specialists in Rwanda is a major challenge with skills concentrated mostly at the National University of Rwanda (NUR) Centre for GIS (CGIS) due to the project (2005–2008) funded by NUFFIC (Netherlands Program for the Institutional Strengthening of Post-secondary Education and Training Capacity). CGIS cooperated with the Faculty of Geoinformation and Earth Observation (ITC), University of Twente, the Netherlands to develop GISc based modules. These modules were incorporated in existing undergraduate and postgraduate programmes and new postgraduate programs in GISc, environment and sustainable development as well as land administration were also developed. Overall, 1 PhD, 14 Masters candidates and 22 short course graduates were trained.

For projects to succeed when qualified personnel are lacking, the initial step is to focus on building human capacity in order to develop a critical mass of geospatial specialists (technical and professional). In the new economy, the capability of organizations to radically innovate, create and sustain competitive advantage has moved from the tangible to intangible assets which reside in human resource (Kessels 2009).

Developing human resources requires investment in personnel and infrastructure. One strategy used is to organise refresher courses with the aim of updating participants' Geo-ICT knowledge (Table 1). Table 1 shows some examples of

Table 1 Selected refresher short training conducted locally

Title	Date	Participants	Institution
Putting health on the map: Addressing public health challenges using spatial data and geo-information tools	12–23 Nov. 2012	Rwanda (Rw), Tanzania, Burkina Faso, Ethiopia, Kenya, Ghana, Malawi	CGIS, ITC, Dutch Royal Tropical Institute
Developments in water quality monitoring, optimization and modernization of surface water quality monitoring programs	22–26 Oct. 2012	NA	NUR, UNESCO-IHE http://www.unesco-ihe.org/content/view/full/6166
What's new in ArcGIS 10?	15 Dec. 2011	Rw	ESRI Rwanda Ltd
Introduction to GIS	18–22 Apr. 2011	KIST staff	KIST, ESRI Rwanda Ltd
ArcGIS 10	13 Oct. 2010	NUR staff	CGIS
Introduction to GIS	7 Jan. 2010	SOPYRWA staff	CGIS
Use of low cost earth observation data in environmental & climate monitoring applications	5–16 Oct. 2009	30 participants from Rw, Nigeria, etc.	CGIS, ITC
Geoportal development	3–7 Aug. 2009	Rw, Democratic republic of Congo (DRC), Uganda	Nature Conservancy Costa Rica, CGIS, Southern Missisipi University, USA
Ecological gap analysis	10–14 Aug. 2009		
Installation and configuration of a Geonetcast ground receiving station	6–10 April 2009	NA	Regional Centre for Mapping of Resources for Development, Kenya, ITC
GIS/freshwater biodiversity training	11–15 Dec. 2006	17 participants from Burundi, Rw and DRC	IUCN, CGIS, ITC
Geo-informatics for hydrological modeling	3–4 Oct. 2005	10 African countries	ITC, UNESCO-IHE, CGIS

Source Author's compilation; *NA* not available

refresher courses organised locally. For details of some GI short courses now available locally, see Akinyemi (2012).

Another strategy is the GIS in Rwanda secondary schools (K-12) launched in 2007 and facilitated by the ArcGIS software donation from Jack Dangermond (CEO ESRI Inc.) in 2006. It incorporates GIS in schools ICT curriculum, ESRI summer camp is organised, teachers undergo GIS training, and a Rwandan GIS textbook was produced. In schools, the focus is not so much on teaching GIS, as it is on GIS use to teach the existing curriculum. Many teachers have their students do exercises in environmental sciences using GIS as a tool for the analysis of the data collected (Phoenix 2004). This initiative is a partnership between government

(Ministry of Education, CGIS implementing) and the private sector (ESRI Inc., ESRI Deutschland GmbH and ESRI Rwanda Ltd) (Forster et al. 2007; Forster and Mutsindashyaka 2008; Huynh 2009).

The ESRI summer camp is organised annually with partners. In 2009, the camp was conducted on environmental issues on Lake Muhazi (http://www.esri.rw/news/articles/n091218.html). In 2011, students collected GPS points of touristic facilities in Kigali (http://www.esri.rw/news/articles/n111213.html). Through this initiative, youths are getting interested in GIS and a critical mass of youngsters with basic GIS knowledge is being produced. Invariably some of them will go on for further studies in GISc.

4 Overview of a GISc Program

As an example of a GISc program, the Applied Geo-Information Science Post-graduate Diploma/Certificate (PGD/PGC) Programme is examined (Table 2). It offers introductory, advanced and application courses.

PGD participants select either an environmental track or urban/society track and do application courses according to their preferences. The combination of

Table 2 Overview of the courses offered in the applied GISc programme

	Credits	Total
Introductory courses		
Introduction to GIS	12	
Introduction to Remote Sensing	12	
Cartography and Visualization	8	
Spatial Data Base Management Systems	8	40
Advanced courses		
Advanced GIS	8	
Advanced Remote Sensing	8	
GISc for Multicriteria Evaluation Techniques	12	
Advanced Cartography/Web mapping	8	
Spatial Statistics	8	44
Application courses		
Environmental track		
Environmental Modelling with GIS and RS	12	
GISc for Water and Wetland Management	12	
GISc for Environmental Impact Assessment	8	
Urban/society track		
GISc for Land Administration	12	
GISc for Urban and Regional Planning	12	
GISc for Land Evaluation/Suitability Analysis	12	
Individual Case Study	16	84
Total		168

Source CGIS-NUR (2008)

these courses can be used to compile different trajectories for a professional clientele, for GIS users with more complex application problems and to provide tailor-made training solutions for institutional clients. These courses can also be used for non-credit training (certificate of attendance) (CGIS-NUR 2008). In summary, this programme is interdisciplinarity in nature, provides a flexible and streamlined curriculum leading to practice-oriented specializations in GISc.

Series of workshops and meetings were held, particularly with stakeholders, to ensure they patronize the program. A training need assessment was conducted which revealed that a considerable number of institutions in Rwanda are aware of the potential benefits of GIS/RS. It also showed the prime application areas for GIS/RS and a preference for training of relatively short duration. Initially, it was designed as a full-time day program where modules are to be taught in an intensive manner in blocks splitted into several courses of between 2 and 3 weeks duration. However, in 2009, at its commencement, most organizations wanted an evening part-time program in Kigali (the capital city) that will allow employees to study and continue working. Thus, the program structure was amended to accommodate this request. This goes to show the need for stakeholders input in program design and delivery (DiBiase 2004).

5 Discussion and Conclusion

GI human capacity in Rwanda is developed locally through formal educational programs, refresher courses, tailor-made courses and K-12. These evidently equip personnel with needed skills to meet the increasing demand for a geospatial workforce (Gewin 2004; Phoenix 2000). Satisfying increasing demand for GI education in Rwanda is challenging as academic staff turnover is high. This threatens stability in faculty positions in existing GI programs. Also with educational budgets reducing, it's challenging to mount new programs and expand existing ones. GI educators must consider future trends, and how their own programs need to be ahead of the theoretical/technological curve (McMaster and McMaster 2012).

5.1 Market Need for GI Skills

A clearer understanding of marketplace needs and demands as well as the advantages and disadvantages of different delivery models for GIS education are required. Looking at market demand in Rwanda, the areas of need is for skilled manpower in both core and applied GI. The latter is drawn from the graduate pool within disciplinary enclaves but with GIS knowledge, e.g. biology conservation, agriculture. Whereas, the former is drawn from among graduates of geomatics, geography and computer science. GI educators must proactively respond to market

demands in a quickly changing profession (McMaster and McMaster 2012). How well does existing courses satisfy market need? Are there emerging areas to be added? How well do the curricula fit, in particular, with the GISc and Technology Body of Knowledge (UCGIS 2006). GI educators must consider future trends, and how their own programs need to be ahead of the theoretical/technological curve. There is the need to consider professional students who are retooling into GIS from other areas such as engineering or urban planning.

Examining available GISc programs in Rwanda reveals the lack of modules on location based services/mobile technologies and web-based access methods (Akinyemi 2012). More computer based modules are needed to enhance student ability for programming as organizations sometimes develop and/or tweak existing software to implement their own solutions. Also GI graduates need skills in business and management, particularly entrepreneurship. Our professionals must understand e-commerce and marketing principles and satisfy client needs. Furthermore, legal aspects of information systems and data are major components of the business pillar. Graduates must know the legal aspects (e.g. privacy rights, data copyright, legal liability), legal impacts on the use of databases and spatial datasets, as well as legal options to deal with conflicts (Onsrud and Rushton 1995; Frank and Raubal 2001).

5.2 More Educational and Training Options

McMaster and McMaster (2012) advocated for developing effective models of access given the growing demand for GI education worldwide. In Rwanda, entry into existing programs is very keen with only the very best students getting admission with limited government sponsorship. To ensure high quality and standard in course delivery, relatively good student–lecturer ratio, student–computer ratio must be maintained.

The typical model of program delivery in use is the residential model. According to McMaster and McMaster (2012), this model assumes students will succeed through being on campus and undertaking formal lectures, working in laboratories with other students, and benefiting from other types of formal and informal professional and social interactions. Students also have easy access to campus resources and more opportunities for face-to-face interaction with faculty, students, staff and local professionals (especially if the institution is located in a metropolitan area). Interaction with faculty expertise and advising is theoretically easier, students do not have to purchase their own software, and access to libraries and other resources is less complicated. On-campus programs can also be beneficial since many successful programs have a dedicated staff person who assists with the day-to-day management of the program including handling student needs in a timely manner. A major disadvantage is that students must commit to relocating to a campus. For professionals who may already have an established career along with other personal commitments it is extremely difficult to relocate. Also,

courses are offered on a fixed schedule that provides little flexibility for the working professional GIS student but some key courses can accommodate professional students by being offered in the evening.

In addition to the traditional course delivery model, other approaches that are yet to be explored in Rwanda are online GISc degree certification and distance learning. Online certificate programs and masters degrees are easily found with a simple Google search on the words GIS and education (Phoenix 2004; Brox et al. 2006). Another option is Web-casting of GI courses which entail broadcasting existing GISc courses across the web using video streaming of class sessions in combination with web hosting of assignments, lecture outlines and reading materials (Onsrud (undated). Some institutions such as NUR and Kigali Institute of Science and Technology (KIST) already have iLab facilities (http://icampus.mit.edu/projects/ilabs/).

Another distinguishing characteristic of professions is specialized certification or licensure. We typically think of these as mechanisms to ensure that individual practitioners are competent and trustworthy. However, another way to think about certification is as a road map for continuing professional development. Increasingly, practicing professionals in different professions are being licensed and certified before they can practice in Rwanda. The GIS discipline is not different but as a rapidly changing field where the technology is being adopted by all types of businesses and professionals, flexibility is required in defining who is a GIS professional (Cotney 2012). The current trend is to have professional certification, for example, the 5024 individuals from 35 countries that are certified as GISPs (GIS professionals) by the US based GIS Certification Institute as of year 2012.

5.3 Recommendations

To approve educational programs in Rwanda, it is mandatory that such programs are in areas of critical need, contributing to developing needed manpower to achieve the Vision 2020 (MINECOFIN 2012). It is essential that programs are sustainable, that is, largely self sustaining financially. The following strategies are recommended to further develop geospatial manpower in Rwanda:

- Stakeholders' need assessment and input before the program commences guarantees their patronage (sponsoring employees) to ensure program sustainability. Private–Public partnership is required where organizations advocate and support program development.
- Employees on part-time training are allowed to close early to attend evening classes daily, thus supporting employees to continue working while studying.
- With expertise in a critical area, GI professionals could commit certain hours weekly to training and teaching.
- Including an industrial attachment component in existing GISc programs to ensure graduates are conversant with the demands of the industry.

References

Akinyemi FO, Uwayezu E (2011) An assessment of the current state of spatial data sharing in Rwanda. Int J Spatial Data Infrastruct Res 6:365–387

Akinyemi FO (2012) Mainstreaming the use of geographic information in a developing country context: examples from Rwanda. In: Proceedings, Geo-Information Society of South Africa (GISSA) conference, 3–5 Oct 2012, Guateng, South Africa. www.eepublishers.co.za/images/upload/GISSA%20Ukubuzana%202012/Felicia_Akinyemi.pdf

Brox C, Riedemann C, Kuhn W (2006) Exchange of complete e-Learning courses—first experiences with a business model. In: Proceedings, European GIS education seminar, September 7–10, 2006, Krakow, Poland

CGIS-NUR (2008) Study guide for the postgraduate short course certificate/diploma programme in Applied Geo-Information Science

Cotney K (2012) GISP and flexibility of the EDU requirement. GIS Certification Institute Newsletter, November 9 issue, 2012

DiBiase D (2004) Engaging stakeholders in program planning for an online master of GIS degree program. In: Proceedings, annual conference of the American congress on surveying and mapping

DiBiase D (2012) Strengthening the GIS profession. ArcNews. Summer 2012. http://www.esri.com/news/arcnews/summer12articles/strengthening-the-gis-profession.html. Accessed 16 Nov 2012

Forster M, Schilling M, McConnell T (2007) Introducing GIS to K12 education in Rwanda. In: Proceedings, ESRI Education and User Conference. http://proceedings.esri.com/library/userconf/educ07/educ/papers/pap_2041.pdf (Accessed 13 Feb. 2012)

Forster M, Mutsindashyaka T (2008) Experiences from Rwandan secondary schools using GIS. In: Proceedings, ESRI education and user conference. http://proceedings.esri.com/library/userconf/educ08/educ/papers/pap_1119.pdf. Accessed 13 Feb 2012

Foundyller C (2011) Geospatial sales strengthened in 2010, growth expected to continue. Waterworld. http://www.waterworld.com/articles/2011/01/geospatial-sales-growth-expected.html. Accessed 12 Nov 2012

Frank AU, Raubal M (2001) GIS education today: from GI science to GI engineering. URISA J 13(2):5–10

Gewin V (2004) Mapping opportunities. Nature 427(22):376–377

Gondo T, Zibabgwe S (2010) GIS solutions and land management in urban Ethiopia. Perspectives on capacity, utilization and transformative possibilities. Manag Res Pract 2(2):200–216

Government of India (2010) Human resource development strategies for Indian renewable energy sector. Ministry of New and Renewable Energy and Confederation of Indian Industry. http://mnre.gov.in/file-manager/UserFiles/MNRE_HRD_Report.pdf. Accessed 13 Nov 2012

Huynh NT (2009) The role of geospatial thinking and geographic skills in effective problem solving with GIS: K-16 education. Theses paper 1078, Wilfrid Laurier University

Kessels WMJ (2009) Designing favourable learning environments in an emerging knowledge economy. In: Workshop presentation at the Kigali Institute of Management, 31 August, 2009, Kigali, Rwanda

McMaster SA, McMaster RB (2012) University of Minnesota master of GIS (MGIS) program a decade of experience in professional education. In: Unwin DJ, Foote KE, Tate NJ, DiBiase D (eds) Teaching GIS and technology in higher education. Wiley, Hoboken

MINECOFIN (2012) Guidelines for development of sector strategies in the context of *Economic Development and Poverty Reduction Strategy* 2 elaboration. GoR, Kigali

Onsrud HJ, Rushton G (1995) Sharing geographic information: an introduction. In: Onsrud HJ, Rushton G (eds) Sharing geographic information. New Jersey: Centre for Urban Policy Research. New Brunswick, pp 8–18

Onsrud HJ (undated) Web-casting of geographic information science graduate courses. http://www.spatial.maine.edu/~onsrud/pubs/WebcastingOfGIScienceGrdCrs.pdf. Accessed 16 Nov 2012

Phoenix M (2000) Geography and the demand for GIS education. Association of American Geographers (AAG) Newsletter, June, p 13

Phoenix M (2004) GIS education: a global summary. In: Proceedings, 4th European GIS education seminar, 2–5 September 2004, Villach, Austria

Ramli R, Noah SA, Yusof MM (2010) Ontological-based model for human resource decision support system. Lecture notes in computer science, Springer-Verlag, Berlin, Heidelberg, vol 6428, pp 585–594

Republic of Uganda (2010) National development plan (2010/11–2014/15)

Schilling M, Twarabameneye E, de Vries W (2005) Geo-ICT capacity building in Rwanda. GIM Int 19:6

Technavio (2011) GIS market in retail sector 2010–2014, p 27. http://www.reportsnreports.com/reports/75993-geographic-information-system-gis-market-in-retail-sector-2010-.html. Accessed 12 Nov 2012

Technavio (2012) Global GISs market 2011–2015, p 39. http://www.technavio.com/content/global-geographical-information-systems-market-2011-2015. Accessed 12 Nov 2012

UCGIS (2006) GIS & body of knowledge. AAG, Washington

Issues in Cartographic Education: How and How Many?

David Fairbairn

Abstract This chapter addresses two related issues in contemporary cartographic education: how we should teach and train the increasingly broad range of subject matter which characterises the discipline of cartography today (i.e. what is the scope of cartographic education); and how many should be taught to be cartographers (i.e. what is the scale of cartographic education). A personal interpretation is attempted of these two issues and it is concluded that a well-researched 'Body of Knowledge' can assist in determining the scope of education; and that significantly more effort needs to be directed towards manpower planning in and for the cartographic industry to ensure a meaningful focus for such educational provision.

Keywords Cartographic education · Manpower planning

1 What is a Cartographer?

Like the definition of cartography, and indeed the definition of cartography's key instrument, the map, the meaning of the word 'cartographer' is subject to broad interpretation. What can be agreed upon, however, are some common characteristics of a cartographer's work, and hence some of the aspects and issues which need to be instilled in apprentice cartographers as they develop knowledge and skills through education and training. Thus, a cartographer must possess attention to detail; understand the transformations inherent in the mapping process; have a comprehensive view of the world and the complex processes which occur in it; be knowledgeable about the geospatial datasets which reflect that complexity, datasets which are sourced from the world and are used to represent it; understand the

D. Fairbairn (✉)
School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK
e-mail: dave.fairbairn@ncl.ac.uk

possibilities and limitations of using such datasets for scaling, visualising, archiving, analysis and decision making; ensure communication of information through a unique medium (the map); show an ability to manipulate and process data whilst retaining accuracy; and create synergy of information within an aesthetic framework (Fairbairn 2012).

To illustrate the broad scope to subject of cartography, the following limited examples are presented as typical tasks undertaken by someone who could realistically call themselves a 'cartographer':

- Studying and applying the creativity required to undertake effective symbol design for maps;
- Examining the impact of specific scientific procedures (take, as an example, Douglas–Peucker line generalisation) on spatial data and map features;
- Mastering the esoteric computer systems skills needed to successfully set up a server which can respond to a web map service (WMS) request for a map layer;
- Developing a system for the mass digitisation, recording and dissemination of map documents in a reference library.

These tasks are indicative of a wide-ranging discipline, which is becoming broader every day. Such breadth must be addressed in the educational syllabuses which are designed to meet the interests of trainee cartographers and those organisations engaged in undertaking tasks such as these.

2 What Does Cartography Cover?

Unfortunately, an ever-widening syllabus attempting to address this breadth will spread the educational effort evermore thinly, and lead to unwelcome effects. Covering the full range of the complexity of contemporary cartography runs the risk of being seen as merely conveying specific knowledge on a vast array of topics in just enough depth to allow students to be able to undertake some activities by rote. For the tasks above, the four putative cartographers could be:

- shown Bertin's visual variables, so they know how to create a legend of different point symbols;
- taught how the Douglas–Peucker algorithm works, so they can choose the Point Remove option in ArcGIS generalisation tools to generalise a layer;
- informed that setting up a server is difficult, but told which scripts to invoke to set up a proprietary map server, which they know will be able to respond to a WMS call;
- introduced to an already installed and set up system, ready to have buttons pressed to invoke batch scanning.

It can be seen that this 'learning by rote' does not address the more generic skills and knowledge required of the cartographer, and will result in an inflexible

and limited approach to the acquisition and application of cartographic education. What we should want our cartographers to be able to do, in addition, might be to:

- explore symbolisation in a different way—does Bertin give as much guidance on pictographic symbols as he does on geometric symbols, and if so, can the design be radically changed to these to be even more effective and still abide by his rules?
- understand the alternatives to the PointRemove algorithm in ArcGIS (how does BendSimplify differ?), and adjust the generalisation iteratively and logically by applying different parameter values;
- have the confidence to know what to do when, for example, the installation of Geoserver fails and the linkage with a PostGIS database is lost;
- write a script developing a user-friendly interface which can cope with the effective digitisation of a range of original map products with varying quality.

The recognition of a need for cartographic education to address these much more wide-ranging and flexible requirements to perform tasks contrasts with the cartographic education of previous years. This was an education which presented a linear, fixed set of operations—e.g. compilation, design, drafting and reproduction; or digitising, structuring, layering and digital mapping—which were summarised by the type of flow diagrams popular in cartographic textbooks of the 1970s and 1980s (e.g. Keates 1973, 1989). Today, there are many alternatives to each step in any cartographic flowline, and indeed the flowline is no longer linear—it can branch, backtrack, circumvent; steps can be followed in very different ways using different procedures; and can be skipped entirely. This variety, this availability of data, tools and procedures, and this freedom to select and apply them, is what now characterises cartography and must be reflected in the contemporary education of cartographers.

3 Lighting Candles or Filling Buckets?

It could be suggested that the only way to address the breadth and flexibility required in cartographic education programmes is to follow a distinctly 'liberal arts' approach. Pedagogically, this would initiate a process of contemplation of possibilities and opportunities described in general terms. Students would absorb these, then look inward to their own inner imagination and resources, and ensure they then blossom with understanding and perceptiveness, revealing to the world the inherent nature of their discipline, the value of their self-driven intellectual journey and their ability to *illuminate* whatever challenges they are faced with. The students of such programmes become, in effect, 'candles', burning with an inner confidence about their abilities to approach a range of tasks and circumstances, and *solve problems*—but perhaps with limited technical skill to do so.

Such an approach to 'education by osmosis' could be regarded as highly idealistic; this is an approach which does not recognise the reality that, in cartography, there *are* truisms to be absorbed, there *are* facts to be imparted, there *are*

procedures to be learnt, and there *are* principles to be acknowledged. The teaching of any technological discipline requires the student to accumulate facts and knowledge, store these and then apply them according to fixed norms in order to reach solutions. This form of technological education, contrasting to the 'candle lighting metaphor', could be regarded as 'bucket-filling'.

As the ICA definition of 'cartography' points outs, however, *technology* is only one of three frameworks within which the subject is bound. The *scientific* flavour of cartography must be taught also, and should use the conventional paradigm: tasks of observation, synthesis, testing, modelling, presentation, archiving, and dissemination are effective and accepted methods of addressing the fundamental transformations inherent in the mapping process; whilst the *artistic* angle can be incorporated into cartographic education to promote a connection with the real world through encouraging creativity, strengthening spatial knowledge, and training in visual thinking. Such 'lighting of candles' must supplement and enhance the 'filling of buckets'.

4 Contemporary Pedagogy

So how does contemporary cartographic education address these contrasting objectives of encouraging innovative flexibility, use of the scientific method, development of creativity, and strengthening of the basic principles? Educators in cartography are no different to all other cartographers—modern developments have been embraced with enthusiasm and, as a result, cartographic education has been examined, moderated and modified significantly in recent years.

From a general educational perspective, change has been driven by possibilities of modifying learning objectives, and the seeking of new methods of engagement with the end recipients (the students). Education is often characterised today as a collaborative exercise involving a common journey undertaken by teacher and student together, with common goals of investigation, analysis, synthesis and the achieving of understanding. Problem-solving approaches, student-centred learning, the development of e-portfolios, the building of knowledge through team discussion, the use of teachers as 'mentors' rather than 'supervisors', and the presentation of an interactive experience to the student: these are all ways of putting the student at the heart of the learning process, making them more responsible for their learning and thus engaging them with the process more, eventually moving away from bucket-filling towards attempts to light candles. There are clearly implications for curriculum design and teaching practice resulting from such approaches (see, for example, Schultz 2012).

Contemporary trends in pedagogical research associated with such developments in educational provision stress the importance of hands-on, practical, 'doing' exercises; learning by exposure to examples, good and bad; and the regular usage of tools and systems to ease the interaction between student and learning material. In cartography, such activities can be assisted by recent developments in

interactive tools to assist in many stages of map production, for example Bernhard Jenny's FlexProjector (Jenny et al. 2010); the familiar colorbrewer, created by Mark Harrower and Cindy Brewer and now embedded in a number of GIS software packages (Harrower and Brewer 2003); Ben Sheesley's developing typebrewer (Sheesley 2007); and Olaf Schnabel's map symbol designer (Schnabel 2005). Provided they are used in a supportive and interactive environment, such cartographic tools are all exciting and useful pedagogical tools for improving student awareness, imagination, and understanding. Cartographic education has rightly taken full benefit from these, and other, resource-driven initiatives.

5 Educational Programmes and Content

In a wider context, and following the points raised above, a contemporary approach to the challenge of developing cartographic education needs to cover the broad scope of the subject, and fulfill the need for both 'bucket filling' and 'candle lighting'. Such an approach can be assisted by the creation of a 'Body of Knowledge' relevant to geospatial science.

Starting with a broad approach to address the core knowledge in Geographic Information Science and Technology (GIST), the Body of Knowledge published under that title (after a decade of work) in 2006 by the American Association of Geographers (DiBiase et al. 2006) did make some acknowledgement of cartography's special role. The broad knowledge area, 'Cartography & visualisation' included the themes 'History & trends', 'Data considerations', 'Graphic representation techniques', 'Map production', and 'Map use & evaluation', each further divided into a number of topics. The majority of the other nine GIST knowledge areas had some further themes and topics which might be regarded as 'cartographic' in nature. At the topic level, educational objectives were presented for each, and from these a coherent Model Curriculum was developed, indicating what geospatial professionals should know and be able to do.

The merits of this approach to a Body of Knowledge for cartography are that it is interdisciplinary (cartography is shown to be intimately connected to other disciplines), it is pedagogically sound (it lists learning outcomes / educational objectives which can be assessed), it links to research (advanced topics and progress in the discipline are incorporated), and it sets the discipline in the public arena (the impact of, and on, human society is highlighted: this is not just a technological checklist).

6 Developing the Body of Knowledge

Current developments in updating the Body of Knowledge (which now includes some European input to what was originally a US-led initiative) include a

continued recognition of the importance of human interaction in cartography, along with further enhancement and addition of uniquely cartographic aspects. These are of interest to ICA, its Executive Committee and Commissions, and have a potential impact on its Research Agenda, and its mission. Five core areas have been identified, most notably by the ICA leadership, each technologically defined, but each of which also involves the discipline of cartography engaging with its wider community and the general public: Data acquisition and Sensor networks; Internet cartography, Web Mapping and Social Networks; Location Based Services, Ubiquitous Computing and Real-time cartography; 3D, Augmented Reality and Cross Media; and Geospatial data infrastructures (Gartner and Schmidt 2010). Flavoured with more theoretical concepts, such as neo-geography, and an appreciation that the basic task of cartography remains unchanged, it is to be hoped that this ICA-led input to the Body of Knowledge can assist in developing contemporary Model Curricula, with a foundation of cartographic fundamentals, suitable to address the objectives of cartographic education presented above.

Although there is acknowledgement within the Body of Knowledge that there are different and progressive intellectual levels of exposure to, and achievement with, the various topics presented, there is little recognition of the significant variability in educational activity which occurs throughout formal and informal educational systems. It is important to recognise, for example, that education in cartography can take place within differing paradigms of general education (for example, a crowded classroom requires different educational approaches to a one-to-one tutorial); different age groups can be involved—education of children is different to education of experienced managers; educational establishments differ in approach and purpose (a university is different to a technical college, and different to a primary school); differing pedagogical models (for example, for distance learning, for student-centred learning, or for continuing education), exhibit variability.

Such variability is not recognised in the surveys of educational courses in cartography by the ICA Commission on Education and Training (http://lazarus.elte.hu/cet/undergraduate/index2012.htm), which concentrate on worldwide provision in one of these areas—full-time education at university undergraduate level. In many parts of the world, there is a greater demand for education and training in cartography than can be met by formal courses such as these, and there is a need to address the differing pedagogical models. And, as attendance at such formal courses becomes ever more expensive (both financially for the students, and in terms of time), and as government investment in well-equipped institutional classrooms (possibly only used for short periods during 40 % of the days of the year) becomes a lower priority, it seems that the answers to the question 'where will cartography be taught?' will become increasing varied. Rather than formal courses, informal workshops will become more common. Such workshops can be and are being offered by learned societies (such as ICA), by pan-governmental outreach programmes (for example, those associated with organisations such as UNECA or the World Bank), by charities (for example, Water Aid in sub-Saharan Africa), and by institutions which have been set up to commit to

extension teaching away from their main base (for example, ITC in the Netherlands). Most of these workshops are not promoted as replacements for formal classes, but as ad hoc attempts to meet rising demands. Commercial companies can also play a large part in delivering such workshops, concentrating notably on their own products and methods. Further, the collaboration between some commercial course suppliers and well-respected educational institutes in developing and delivering on-line courses is becoming increasingly common and effective (e.g. Coursera and Penn State).

On-the-job training is a further similar approach to education outside the classroom. The need for a well-trained workforce is regarded as paramount by most progressive corporations, and with procedures and possibilities in the revised cartographic flowline changing constantly, it is essential that a well-educated workforce can receive updates in the workplace itself, by means of continuing professional development, and by courses and workshops. The Body of Knowledge can assist in the development of learning objectives for both such focussed training courses, and any more general educational provision in the discipline of cartography.

7 Planning

The question explored above of how we teach cartography has an important supplementary—how many do we teach? The answer lies in the perceived demand for trained and educated cartographers. Noteworthy statistics have been presented in the past few years by those who are attempting to gauge those numbers. A striking calculation from Molenaar (2009) estimates that there are currently 2.5–3 million geomaticians employed worldwide: if each has a 40 year career then 75,000 new recruits are needed each year; and if 10 % of the total workforce needs some form of continuing educational provision in order to update, then a further 300,000 students can be counted upon. A significant proportion of these potential geomatics students will be cartographers. Already, skills gaps in geomatics and other similar activities are evident in many mature economies—the only answers to such gaps are to encourage qualified immigrants and/or to improve the employability and skills of the current student body and the existing workforce. The former relies on a supply of educated and ambitious graduates, notably from developing countries, and justifiable accusations can be made of 'creaming off' the best and brightest from such countries for the benefit of developed economies. If we are, instead, to emphasis the latter option, then some estimate of the numbers involved must be sought. This is an exercise which seems to be anathema to most free market nations: the notion of 'Stalinist' centralised manpower planning has no place in most western economies. However, there are interesting and instructive examples of the quantitative assessment of employment patterns and market sectors. The United States Department of Labor has a long-standing interest in reporting on and improving the American labour market, and in 2010 released its

Geospatial Technology Competency Model (accessible through http://www.careeronestop.org/competencymodel/).

Whilst not giving specific instruction about employment rates and numbers within sectors in geospatial technologies, the model is quite prescriptive in its coverage of expected skills, background abilities, and professional standing of expected entrants into the geospatial profession, and thus has an influence on the way in which the workforce in areas such as cartography and other geospatial sciences can be directed and managed.

Further north, a decade ago, the Canadian government, through its Department of Human Resources and Skills Development commissioned an extraordinarily comprehensive Human Resources Study on the Geomatics Sector of the Canadian Economy (CCLS et al. 2001). Prepared by learned and professional organisations in the Canadian geomatics arena, this report covered a wide range of issues, including description of the disciplines within geomatics (including cartography), the nature of the industry (both inside Canada and also how well Canadian geomatics works overseas) and a thorough review of the technology currently used and how it would develop in the future. Two further extensive sections of the report describe education and training in geomatics in Canada, and the human resources profile. A survey of geomatics activities showed expected increases in activity in all sectors in the 2001–2006 period, but some, including photogrammetry, geodesy, cartography and land surveying, were expected to increase at a slower rate than other areas such as navigation, GIS, decision support and consulting. The overall rate of growth led the survey to estimate a need for two thousand university graduates in geomatics by the year 2004, compared to an estimated supply of 950 at the time of survey in 2001. Again, a significant number of these would be cartographers. In retrospect this was an underestimate, and particularly in cartography a more vibrant sector would require an even greater increase in trained personnel during the first decade of this century: the survey was not able to foresee the extraordinary growth in the sub-disciplines of cartography as highlighted by Gartner and Schmidt (2010). The demand for experienced cartographic graduates in internet cartography, web mapping, LBS, SDIs, sensor networks, and augmented reality must be about to increase even more, widening the skills gap in the cartographic industry between demand for trained and educated cartographers and supply of them even further.

The final example of manpower planning and labour market analysis presented here is more contemporary, and does take into account the rapid developments of the cartographic discipline. A Final Report on manpower strategies published by the Botswana Training Authority in 2010 (BOTA 2010) was commissioned to 'forecast and identify a list of priority vocational skills and develop strategies to fast track priority skills development'. Its comprehensive overview of the national economy, the health of market sectors, the basic methodology adopted to data collection and skills forecasting, and the list of critical and priority skills, mark this report out as a document of considerable importance. One particular outcome of interest is that even in a nation with a population as low as Botswana's (just over 2 million in 2011), there is an estimated shortfall of over 4,000 educated and trained

geomaticians by the year 2016. It is incumbent on those who educate cartographers to ensure that provision is made to match such needs throughout the world. Professional organisations in cartography, and large employers of cartographers in industry and government should be urged to produce national overviews of manpower requirements for cartographers, and to ensure that such overviews are translated, through government policy, into practice. The provision and take-up of educational programmes in cartography is dependent on the perception of a need for cartography within a national economy; and there is a corresponding need in industry and government for there to be a supply of trained and educated cartographers to meet that need.

8 Conclusion

There are global challenges in the future which must be met with reliance on the skills, experience and imagination of well-educated cartographers. But there are challenges within cartography itself, two of which, in particular, have been highlighted here in this overview of cartographic education. Firstly, cartographic education must correctly balance education in the basics and the fundamentals, with education for sophistication and enlightenment. Secondly, we must ensure that we are able to educate sufficient numbers of cartographers to ensure that the valuable role of cartography is maintained and expanded, and that the discipline of cartography retains its identity.

References

BOTA (2010) Final report. Botswana Training Authority, Gaborone 224 pp

CCLS (Canadian Council of Land Surveyors), Canadian Institute of Geomatics (CIG), and Geomatics Industry Association of Canada (GIAC) (2001) Geomatics sector, human resources study. Human Resources Development Canada, Ottawa, 372 pp

Di Biase D, Demers M, Johnson A, Kemp K, Taylor-Luck A (eds) (2006) Geographic information science and technology body of knowledge. AAG, Washington

Fairbairn D (2012) Candles for cartography: helping to illuminate our discipline. In: Keynote address at GeoCart2012 (ICA regional symposium for Australia and Oceania), Auckland, New Zealand

Gartner G, Schmidt M (2010) Moderne Kartographie-Technologische Entwicklungen und Implikationen. Kartographische Nachrichten 6:299–305

Harrower M, Brewer C (2003) ColorBrewer.org: an online tool for selecting colour schemes for maps. Cartogr J 40(1):27–37

Jenny B, Patterson T, Hurni L (2010) Graphical design of world map projections. Int J GIS 24(11):1687–1702

Keates J (1973, 1st edn) (1989, 2nd edn) Cartographic design and production. Longman, Harlow

Molenaar M (2009) Cross border education for the global GI-community. In: Keynote speech at the 24th international cartographic conference, Santiago de Chile

Schnabel O (2005) Map symbol brewer—a new approach for a cartographic map symbol generator. In: Proceedings of 22nd international cartographic conference, July 2005, A Coruna, Spain

Schultz R (2012) Active pedagogy leading to deeper learning. In Unwin D, Foote K, Tate N, DiBiase D (eds) Teaching geographic information science and technology in higher education. Wiley, Chichester (Ch 9)

Sheesley B (2007) Typebrewer: design and evaluation of a help tool for selecting map typography. PhD thesis, University of Wisconsin-Madison

The State of GISc Education and SDI Implementation in the SADC Countries: A Comparative Study

Sanet Eksteen and Serena Coetzee

Abstract Most of the countries in the Southern African Development Community (SADC) are poor but rich in various natural and agricultural resources. The vision of SADC is a common economic well-being, improved standards of living and quality of life for all people in the member states. Geographical information science (GISc) plays a vital role in managing the natural and agricultural resources to achieve the vision of SADC. Spatial data infrastructures (SDIs) facilitate access to geographical information and provide policies to manage access to geographical information. These are key success factors for finding sustainable solutions to the challenges of all member states. In order to implement SDIs and to strengthen the use of geographical data in solution finding, these countries need GISc education to train the necessary professionals and scientists. The objective of this study is to investigate if there is a relationship between the availability of GISc education and SDI implementation in the SADC countries. To our knowledge a similar study has not been undertaken. Previous studies have been conducted to determine the current state of GISc education and SDI implementation respectively on the African continent. In this chapter we compare the findings of the two studies to determine if there is a relationship between the availability of GISc education and the state of SDI implementation in SADC countries. The results indicate that SDI implementation in SADC is not influenced by the availability of GISc education.

Keywords GIS education · Spatial data infrastructures · Southern African development community · Geographical information science

S. Eksteen (✉) · S. Coetzee
Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
e-mail: sanet.eksteen@up.ac.za

1 Introduction

The Southern African Development Community (SADC) consists of 15 member states of which most are poor countries but rich in various natural and agricultural resources. The vision of SADC is a common economic well-being, improved standards of living and quality of life for all people in the member states (SADC website 2012). Geographical information science (GISc) plays a vital role in providing answers to economic, environmental and social challenges. Spatial data infrastructures (SDIs) to facilitate access to geographical information and policies to manage access to geographical information are key factors for success to find sustainable solutions to challenges of all the member states (UNECA 2004).

The importance of GISc to find sustainable solutions to these challenges and to manage the resources has been confirmed by many national and international reports (Chien et al. 2012; Toure et al. 2012; Adiat et al. 2012). The United Nations Initiative on Global Geospatial Information Management (UN GGIM) and the Group on Earth Observations (GEO) aim to address some of these challenges through the availability of geographic information and earth observations. The aims of these initiatives to address global challenges together with the above mentioned importance of SDIs confirms the need for GISc scientists and professionals who can maintain and analyse the geographic data. However, a recent study by Eksteen and Coetzee (2012) indicated a worrisome trend that GISc education on the African continent is not readily available. Furthermore, a study by Makanga and Smit (2010) indicated that SDI implementations on the African continent are still in their infancy.

In the light of the above, the objective of this study is to investigate if there is a relationship between the availability of GISc education and SDI implementation in the SADC countries. One expects GISc education capacity to be a precursor for an advanced state of SDI implementation, but is this really the case? To our knowledge a similar study has not been undertaken.

The subsequent sections in this article discuss the objectives of SADC and the important role of spatial information to achieve these objectives. In the third section we discuss the methodology followed in this study. In the fourth section we present an overview of the results of the two previous studies on the current state of GISc education and SDI implementation on the African continent respectively, followed by the discussion of the results of the comparative study for the SADC countries. In the last section we conclude and discuss the implications of the results for GISc and SDI implementation in the SADC countries.

2 The Southern African Development Community

SADC was formed on 17 August 1992 when all heads of states and governments signed the SADC treaty and declaration. Currently SADC consists of 15 member

states, namely Angola, Botswana, Democratic Republic of the Congo (DRC), Lesotho, Madagascar, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia and Zimbabwe. SADC supports various objectives. Objectives relevant to this study include the achievement of development and economic growth, poverty alleviation and regional integration. SADC aims to promote peace and security, self-sustaining development and synergy and collaboration between national and regional strategies and programmes. They aim to achieve sustainable utilization of natural resources and the protection of the environment (SADC website 2012).

The 15 member states have a population of about 250 million with a gross domestic product (GDP) of approximately US$470 billion (SADC Website 2012). The natural resources in the area include gold, platinum and diamonds in South Africa (SA web 2012), copper, lead, tin in Namibia and cobalt, petroleum and silver in the DRC (ISSA 2012). However the SADC states are also plagued by natural disasters such as floods and cyclones in Malawi, Mozambique and Madagascar (OCHA 2012) and droughts, veld fires and animal diseases in South Africa and Botswana (UNjobs 2012).

SADC supports various programmes of which the Agricultural Information Management System (AIMS 2008) plays a vital role. The main objective of AIMS main objective is to provide easy access to information regarding agricultural and natural resources. Activities under this programme include the timely collection of information for the use of early warning systems, vulnerability assessments, food security and the monitoring of weather patterns. A more recent development is the Regional Indicative Strategic Development Plan (RISDP) and the Dar-es-Salaam Extra Ordinary Summit Declaration and Plan of Action on Agriculture and Food Security. Both programmes are in need of spatial information. AIMS is actively involved in the establishment of an integrated database and network for Food, Agriculture and Natural Resources in the SADC region (AIMS website 2008).

3 Methodology

In this study we compared the results of two previous studies to determine if there is a relationship between tertiary GISc education capacity and the state of SDI implementation. The first study by Coetzee and Eksteen (2012) determined the current status of GISc education capacity on the African continent. The second study by Makanga and Smit (2010) determined the state of SDI implementation in Africa. For this study we used only the data relevant to the SADC countries for further analysis and mapping.

Nine of the 15 SADC countries took part in the SDI implementation survey conducted by Makanga and Smit (2010). Even though not all member states took part, the results give a good indication of the current state of SDI implementation in the SADC countries.

4 Previous Studies on GISc Education and SDI Implementation on the African Continent

Recent studies indicated the current state of GISc education capacity and related challenges in South Africa and in the rest of Africa (Coetzee and Eksteen 2012; Eksteen et al. 2012; Marais 2008). Studies on the importance, implementation and assessment of spatial data infrastructures (SDIs) abound, amongst others Grus et al. (2007) and Masser et al. (2008). Makanga and Smit (2010) assessed the state of SDI implementations on the African continent. We could not find research that compares or links the state of SDIs to the availability of tertiary GISc education.

The African continent hosts roughly 600 tertiary education institutions. Only one country does not have a single tertiary education institution, namely the Western Sahara. Nigeria hosts the largest number of tertiary education institutions on the African continent, namely 112 (4icu.org 2012) Despite the large number of tertiary education institutions, they do not feature on international rankings. Most of the top ten ranked African tertiary education institutions are from South Africa (Webometrics 2012).

Coetzee and Eksteen (2012) collected information about the availability of GISc education from the websites of tertiary education institutions listed in the 4 International Colleges & Universities directory. The directory includes worldwide higher education institutions that are officially recognised, licensed or accredited by national or regional bodies, are officially entitled to grant four-year under-graduate degrees and/or postgraduate degrees and provide traditional face-to-face learning facilities, programs and courses (4icu.org 2012). They recorded the highest level at which a GISc or related degree is offered.

The study revealed a number of worrying trends about the availability of GISc education on the African continent: only 14 % of tertiary education institutions offer GISc education in some form (including geographical information systems (GIS) as a module, remote sensing and geomatics); a large number of websites (23 %) were dysfunctional and the relevant information could not be accessed; and GISc as a degree is only offered in 11 of the 54 countries in Africa.

A greater reason for concern is the lack of remote sensing education on the African continent. There is not a single tertiary education institution that offers a degree in remote sensing. Remote sensing is included as a module as part other degree programmes at only 8 % of the tertiary education institutions. This may be an indication that remote sensing is not yet recognized as a science on its own (Coetzee and Eksteen 2012).

The studies by Coetzee and Eksteen (2012) and Eksteen et al. (2012) indicate that South Africa leads GISc education on the African continent.

The importance of SDIs for a country can be seen in the various studies on SDI implementation and SDI readiness that have been done over the last few years (Delgado et al. 2005; Harvey et al. 2012; Makanga and Smit 2010). Hjelmager et al. (2008) describe a SDI as a way to facilitate and coordinate the exchange and sharing of spatial data and services between stakeholders from various levels in the

spatial data community. SDIs hold economic potential by making spatial data available and are therefore a key component in the development of a nation (UNECA 2004).

In light of the abovementioned importance of SDIs, Makanga and Smit (2010) reviewed the status of SDI implementations in Africa. They adopted the INSPIRE State of Play methodology where four indicator categories were used: organisational, funding, legal and technical. A set of 14 more specific indicators was formulated. A questionnaire was created based on this final list of indicators and sent to 269 people in 47 African countries. Makanga and Smit (2010) state that many of the SDI activities on the African continent are informal and may be difficult to find through website searches. The information was compiled to establish scores against the indicators.

Makanga and Smit (2010) concluded that there is a need to speed up the implementation of SDIs in Africa. Two organizations were mentioned as important role players in creating clearinghouses: SADC and the Food Aid Organization's Somalia Water and Land Information (FAO SWALIM). However, they point out that these organizations can make data sets available but cannot enforce or drive the implementation of SDIs in individual countries.

5 Results and Discussion

5.1 GISc in the SADC Countries

The 15 member states of SADC host about 112 tertiary education institutions. The number of tertiary education institutions per country is indicated in Fig. 1.

South Africa hosts the most tertiary education institutions, namely 23, followed by Tanzania (17) and the DRC (16). Three countries, namely Lesotho, Swaziland and Seychelles have only one tertiary education institution each. Figure 2 indicates the number of people at country level per tertiary education institution. This is a rough indicator of the availability of tertiary GISc education in relation to the population of the country.

Mozambique has the highest number of people per tertiary education institution, namely 7.5 million, followed by the DRC with 4.6 million. Mauritius has the least number of people per tertiary education institution with 0.3 million and Namibia the second lowest with 0.7 million people per tertiary education institution.

Figures 3, 4, 5 indicate the offering of GISc, remote sensing and GISc degrees in the SADC countries. The size of the symbol on the map is proportional to the number of tertiary education institutions offering the specific education, while the background colour indicates the number of tertiary education institutions offering the education.
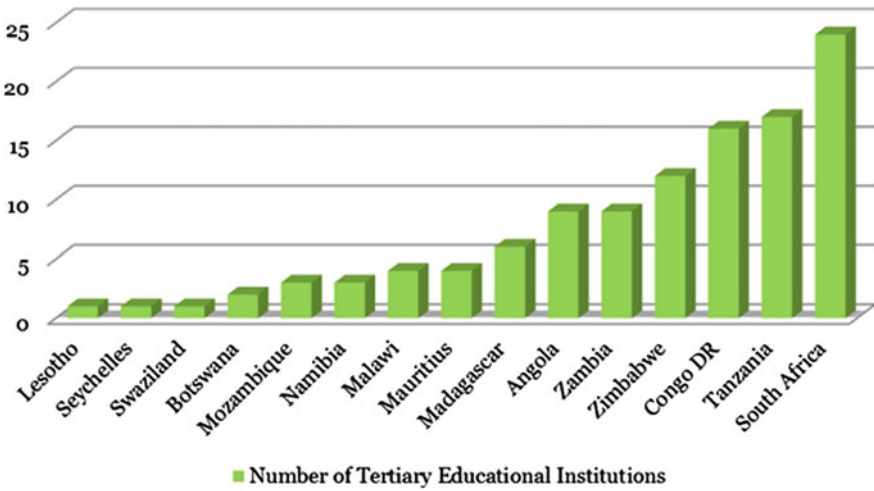
Fig. 1 Number of tertiary education institutions in SADC
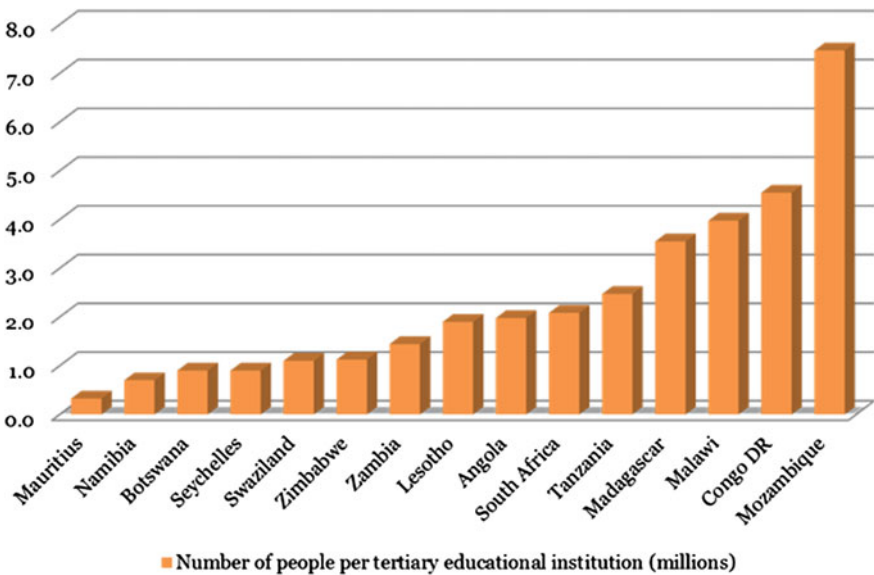

Fig. 2 Number of people at country level per tertiary education institution in the SADC

The offering of GISc education in any form is indicated in Fig. 3. GISc education for the purpose of this study includes GIS, remote sensing and geomatics education. GISc education is offered by only 30 % of the tertiary education institutions in the SADC countries while 13 % of the websites surveyed were

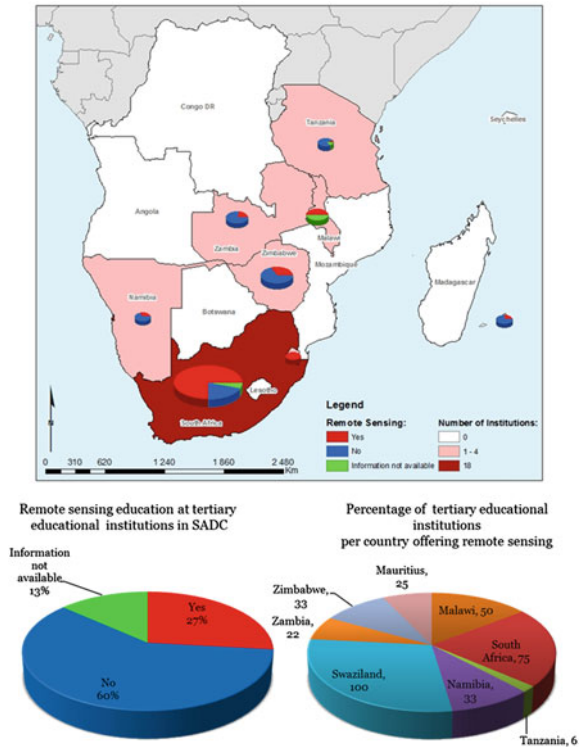Fig. 3 GISc education at tertiary education institutions in the SADC
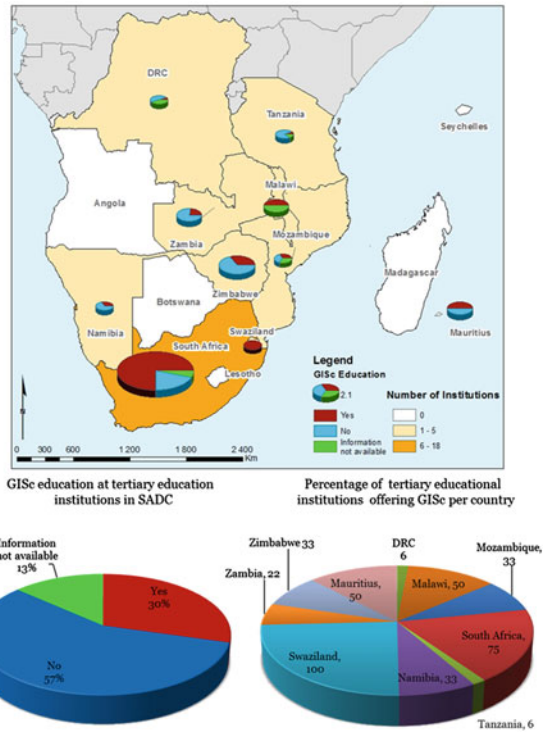
dysfunctional or the information was not available. These percentages are substantially better than those for the whole continent.

The offering of GISc education in the SADC countries is dominated by South Africa. However, the percentage of tertiary education institutions offering GISc education in Swaziland, South Africa, Mauritius and Malawi stands out. A higher percentage of tertiary education institutions offering GISc education results in a higher probability that undergraduate students come into contact with GISc education. Angola, Botswana, Lesotho, Madagascar and Seychelles do not offer any GISc education at all.

The offering of remote sensing is slightly lower (27 %) than GISc education (30 %) in the SADC countries. See Fig. 4.

The offering of remote sensing education is again dominated by South Africa. However, remote sensing is not offered as a degree at any one of the tertiary education institutions in the SADC countries. If one looks at the percentage of tertiary education institutions 0ffering remote sensing as a module, Swaziland (100 %), South Africa (75 %) and Malawi (50 %) dominate the picture once again. However, in Mauritius the offering of remote sensing modules is lower (25 %) than GISc education (50 %). Angola, Botswana, DRC, Lesotho,

Fig. 4 Remote sensing
education at tertiary
education institutions in
SADC

Mozambique and the Seychelles do not offer remote sensing education at any of
their tertiary education institutions.

Figure 5 illustrates the offering of GISc as a degree at tertiary education
institutions in the SADC countries.

GISc degrees are only offered in 4 of the 15 SADC member states. Again the
offering is dominated by South Africa followed by Zimbabwe, Mauritius and
Tanzania.

The offering of Geomatics at tertiary education institutions follows an even
more worrisome trend with only 4 % of the tertiary education institutions in SADC
offering a Geomatics degree. Furthermore, these education institutions are located
in only two of the 15 member states: South Africa and Tanzania. Even the total
offering of GISc and Geomatics degrees at country level is low with degrees
offered in 4 of the 15 countries only.

Fig. 5 GISc degree at tertiary education institutions in the SADC
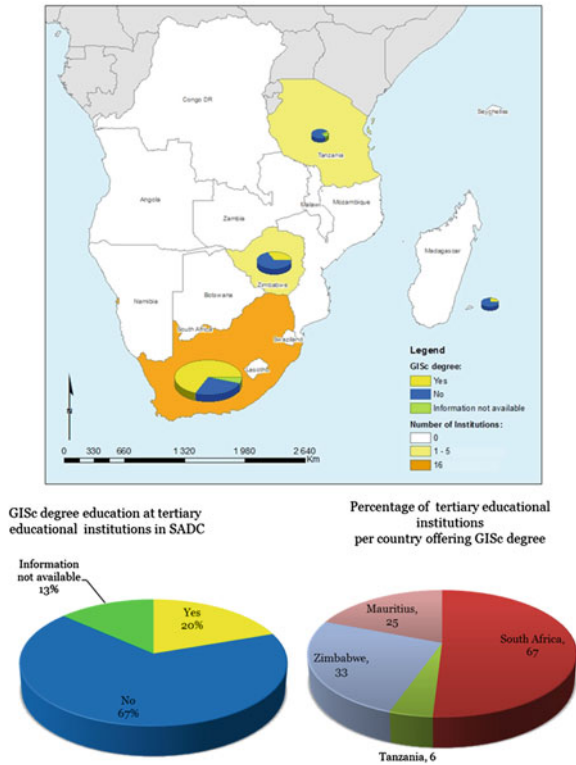
5.2 SDIs in the SADC Countries

The SDI score determined by Makanga and Smit (2010) is summarized in Fig. 6a. They adopted the concept from the INSPIRE State of Play. A numeric value on a scale of zero to four was assigned to responses, where a zero indicated 'Not sure' and a four indicated 'True'. The results were tabulated and totalled to indicate the final SDI score per country. For the purpose of their study they assumed that each of the indicators has an equal weight and that the total indicates the status of SDI implimentation. Tanzania has the highest SDI score, followed by South Africa, Madagascar and Swaziland. Lesotho and Zimbabwe have the lowest scores.

The scores of individual indicators are summarized in Fig. 6b–f. The top three places are dominated by four countries: Tanzania (5 occurrences), Madagascar and Swaziland (3 occurrences), followed by South Africa (2 occurrences). Although South Africa features in only two of the rankings of the indicators, it is listed in the top three SDI scores because of its high score in legal and organizational indicators. Swaziland has a high organisational score, but ends up fourth in the SDI scores because of a relatively low score in the legal and metadata indicators.
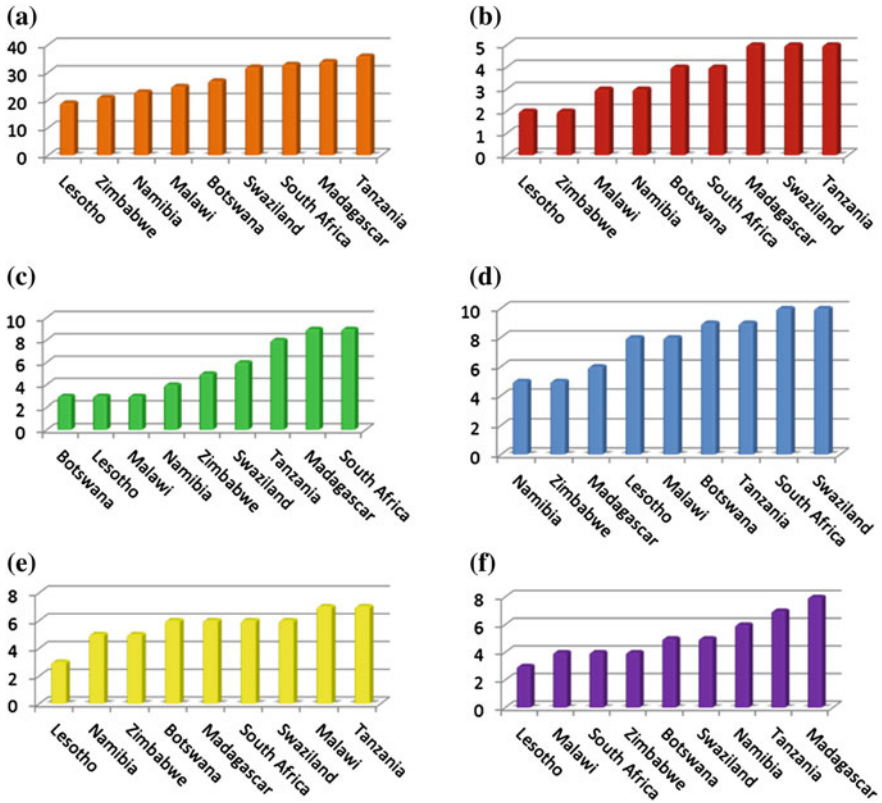
Fig. 6 SDI scores and SDI indicators (Makanga and Smit 2010). **a** SDI score. **b** Funding. **c** Legal. **d** Organisational. **e** Technical. **f** Metadata

6 Conclusion

The results of this research indicate that there is no relationship between the availability of tertiary GISc education and the state of SDI implementations in the SADC countries. Even though one expects tertiary GISc education to be a precursor for advanced SDI implementations, there are countries with a low availability of GISc education and high SDI score and vice versa.

Tertiary GISc education in the SADC countries, though better than in the rest of Africa shows the same worrisome trends indicated by Coetzee and Eksteen (2012) and Eksteen et al. (2012): GISc, remote sensing and geomatics education, similar to the rest of Africa, is not readily available at tertiary education institutions in SADC. GISc education capacity in SADC is dominated by South Africa, followed by Zimbabwe. Swaziland has the highest percentage of tertiary education

institutions offering GISc education, followed by South Africa. Students from these countries have a high probability to come into contact with GISc education.

Although South Africa dominates in terms of GISc education, it does not have the highest SDI score. In contrast to South Africa, Tanzania has the highest SDI score, even though GISc education is not readily available in the country. Zimbabwe follows South Africa when it comes to GISc education but does not feature in the SDI score rankings or any of the individual SDI indicators.

Countries like Angola, Botswana, Lesotho and Seychelles do not offer any GISc education and do not feature in the SDI score rankings or any of the individual indicators. However, Madagascar also does not offer any GISc education at their tertiary education institutions but is listed second on the SDI score. Madagascar's high score is due to its high scores for the legal framework, metadata and funding indicators.

For South Africa, Tanzania and Madagascar the high score for the legal framework has a significant impact on the SDI score. This implies that these countries have the political backing of their governments to implement SDIs. It raises the question posed by Makanga and Smit (2010) if SADC should have the political mandate to implement an integrated SDI for the 15 member states of SADC.

Further research of importance is to determine the reasons why countries with none or only a few tertiary GISc education institutions have a high SDI score. A survey of GISc professionals will be necessary to determine if they obtained their GISc training in other countries. This study focuses on the availability of GISc education; the quality of education and the content of the curriculum are not evaluated. A further study could investigate if the quality of teaching or the course content conveyed to the students influences SDI implementation.

References

4icu.org (2012) 4th international colleges and universities directory 2012, viewed 19 April 2012, www.4icu.org

Adian KAN, Nawawi MNM, Abdullah K (2012) Assesing the accuracy of GIS-based elementary multicreteria decesion analysis as a spatial prediction tool—a case of predicting potential zones of sustainable groundwater resources. J Hydrol 440–441:75–89

AIMS (2008) Agricultural information management systems. http://aims.sadc.int/. Accessed 9 Oct 2012

Chien L, Tseng, W Chang, C, Hsu, C (2012) A study of ocean zoning and sustainable management by GIS in Taiwan. Ocean Coast Manage 69:35–497

Coetzee S, Eksteen S (2012) Tertiary education institutions in Africa: cloudy with a chance of GISc education in some countries. S Afr J Geomat 1(2):119–132

Eksteen S (2012) Coetzee S, Grundling C (2012) Tertiary GISc education at African universities: is the sun shining in South Africa? GISSA Ukubuzana 2012. Ekurhuleni, South Africa, pp 2–4

Fernández T, Kate D, Cuba L, Margaret B (2005) Assessing an SDI readiness index. FIG Working Week

Grus L, Crompvoets J, Bregt AK (2007) Multi-view SDI assessment framework. Int J Spat Data Infrastruct Res 2:33–53

Harvey F, Iwaniak A, Coetzee S and Cooper AK (2012) SDI past, present and future: a review and status assessment. Spatially enabling government, industry and citizens: research and development perspectives. GSDI Association, Needham, pp 23–38

Hjelmager J, Moellering H, Delgado T, Cooper AK, Rajabifard A, Rapant P, Danko D, Huet M, Laurent D, Aalders HJGL, Iwaniak A, Abad P, Düren U, Martynenko A (2008) An initial formal model for spatial data infrastructures. Int J Geog Inform Sci 22(11):1295–1309

ISSA (2012) Institute for security studies. http://www.issafrica.org/. Accessed 17 Oct 2012

Makanga P, Smit J (2010) A Review of the status of spatial data infrastructure implementation in Africa. S Afr Comput J 45

Marais H (2008) The challenges of GIS education and training: GIS use municipal urban and regional planning. Asian conference on remote sensing. http://www.a–a-rs.org/acrs/proceeding/ACRS2008/Papers/TS%2024.1.pdf. Accessed 20 Oct 2012, pp 1–19

Masser I, Rajabifard A, Williamson I (2008) Spatially enabling governments through SDI implementation. Int J Geog Inform Sci 22(1):5–20

OCHA (2012) United nations office for the coordination of Humanitarian affairs. http://ochaonline.un.org/rosa/HumanitarianSituations/tabid/3459/language/en-US/Default.aspx. Accessed 17 Oct 2012

SADC (2012) Southern Africa development community. http://www.sadc.int/. Accessed 10 Oct 2012

SA Web (2009–2011) South Africa web—all about South Africa. http://www.southafricaweb.co.za/. Accessed 17 Oct 2012

Toure N, Kane A, Noel JF, Turmine V, Nedeff V, Lazar G (2012) Water-poverty relationships in the coastal town of Mbour (Senegal): relevance of GIS for decision support. Int J Appl Earth Obs Geoinf 14(1):33–39

UNECA (2004) SDI Africa: an implementation guide. http://geoinfo.uneca.org/sdiafrica/default1.htm. Accessed 12 Oct 2012

UNjobs (2012) UNJobs a Swiss association. Botswana country report submitted to the African regional consultations on disaster reduction workshop, 2–3 June 2004, Midrand, South Africa. http://www.unisdr.org/2005/mdgs-drr/national-reports/Botswana-report.pdf. Accessed 17 Oct 2012

Webometrics (2012) viewed 19 April 2012. http://www.webometrics.info

New Technologies as Educational Resources for Teaching Cartography: A Case Study in Guinea-Bissau

Inês Mario Nosoline, Angelica C. Di Maio and Dalto Domingos Rodrigues

Abstract This work aimed to explore and measure the effectiveness of using geotechnologies to instigate students of elementary and secondary schools, in Guinea-Bissau, to learn more about the issues related to spatial representation. An educational methodology was developed and evaluated in order to provide educators and students with access to digital maps and satellites images. As part of the methodology, questionnaires were applied to teachers with the purpose of identifying and selecting the subjects that were part of the digital educational modules and the content of the databases to be used in Terraview GIS. To evaluate the instructional materials produced, the methodology was applied in four schools including an institute for teachers. The results pointed to the benefits of using new technologies as auxiliaries tools to traditional teaching, the insertion of geotechnologies in the schools activities has facilitated the understanding of the studied subjects, scale and geographic coordinates, thus providing a significant gain in students' performance, which also contributed to the process of digital inclusion and in reducing the lack of teaching materials in Guinea-Bissau.

Keywords New technologies in education · Cartography lessons · Geotechnology in the school

I. M. Nosoline · D. D. Rodrigues
Universidade Federal de Viçosa, Viçosa, Brazil

A. C. Di Maio (✉)
Universidade Federal Fluminense, Niterói, Brazil
e-mail: dimaio@vm.uff.br

1 Introduction

The search for new teaching methods that includes technological advances experienced by society is ever more present when the subject is to encourage the student to create critical thinking and increase the desire to learn. These incentives can be worked out with the collaboration of geotechnology in the school environment, providing advantages as the ability to exploit the students' spatial vision, and in view of the interactivity it provides, it can also help in the process of enhancing the association of school learning with world phenomena. However, the insertion of new technologies in the classroom was only possible due to the ease of obtaining maps and satellite images in digital format via web, and public and private initiatives to provide free GIS (Geographic Information Systems) tools.

In the past, according to Hodgkiss (1981), maps were used for the description of faraway places, as a navigation aid and military strategies resource; it was very difficult to get copies. Today, digital cartography can count with technological advances and developments as the GIS, which led to changes in the speed of production, updating and analyzing geospatial information, and facilitated the combination of information from multi-source spatial data (Bonham-Carter 1994).

These characteristics are especially important in geography studies, as spatial representation contributes to build on the notions of spatial organization, to raise issues and to understand proposed solutions to geographic problems. Such content can benefit from the cartographic language, which can be approached with the use of products of remote sensing and digital mapping, which according to Neves and Cruz (2007), are effective in teaching-learning process, stimulating and awaking students, at different educational levels, to the importance of geographical knowledge in their daily lives and in the process of citizenship construction.

However, to Katuta (2002), the situation experienced in geography lessons is still the underutilization of the map, as a result of its outdated condition or even the lack of it, thus hindering the learning of cartographic communication language, which involves the interpretation of elements such as legend, scale, geographic coordinate system and projections.

Although public policies encourage the use of science and technology as a means of teaching, in many schools, especially in Portuguese-speaking countries, printed and outdated maps are still the main learning tools for cartography. Therefore, the use of digital material could meet the lack of teaching resources; however, one of the barriers found in the integration of new technology in school is still the lack of infrastructure for computer labs. On the other hand, there are free applications on the web in Portuguese language as TerraView GIS (TerraView 2010).

In all societies, education is seen as a strategy in the fight against poverty and violence, and fundamental to human development and to improve life's quality. In the scenario of Guinea-Bissau it is not different, in view of this, a national plan called "Education for All" was defined in the country as a public policy to achieve seven major development goals for the millennium that consisted of: (1) halving

extreme poverty and hunger; (2) ensure basic education for all; (3) promote gender equality and empower women; (4) reducing child mortality; (5) improving maternal health; (6) combat AIDS, malaria and other diseases; (7) ensure a sustainable environment and reduce by half the proportion of the population still without access to drinking water.

Taking into account the Guinean scene experienced in schools, and the difficulty in acquiring books and maps for geography lessons, this study aimed to contribute with an educational resource for both teachers and students. The developed and tested material addressed the topics of geographical coordinates and scale, pointed by the majority of teachers as difficult assimilation topics by the students and of great importance in studies of spatial representation.

2 Methodology

The project was implemented in 4 schools, 3 lyceums and 1 Teacher's Institute, in the region of Bissau in Guinea-Bissau (Fig. 1), located on the west coast of Africa (latitude: 12°N, longitude: 15°W), with the guidelines obtained from the survey on the subjects in which students frequently present difficulties in assimilation. The application of the methodology in the participant schools were authorized by Guinea-Bissau National Institute for Education Development, from Ministry of Education.

The implementation of the project in Guinea-Bissau was conducted as shown in Table 1.

Fig. 1 Location of the study area

Table 1 Activities developed in the selected schools in Guinea-Bissau

Carried out activities	Schools and participant classes			
	Kwame N'Krumah National lyceum (70 students-Grade 8th)	Agostinho Neto lyceum (72 students-Grade 8th and 9th)	João XXIII lyceum (200 students-Grade 8th)	Institute of Education Tchico Té (85 students-year zero)
Pretest	B2 and B3	B8, C5 and C25	B1, B2, B3, B4, B5 and B6	Classes A and B
Scale module and exercises	B2	B8, C5 and C25	B1, B2, B4, e B5	Classes A and B
Geographic coordinates module and exercises	B2	B8 and C5	B1, B2, B4, and B5	Not applied
Google earth and TerraView activities (complementary activities)	B2	B8	B1, B2, B4, and B5	Not applied
Posttest	B2 and B3	B8 and C5	B2, B3, B4, and B5	Not applied
Questionnaires	B2	B8, C5 and C25	B2, B4, and B5	Not applied

2.1 Development

The project development and application in schools took the following materials and programs: GeoEye satellite image of Guinea-Bissau region[1]; Google Earth and Flash programs and TerraView GIS.

The methodology was divided into two steps, the first step referred to the definition of school grade and themes to be covered. A questionnaire was given to 40 teachers. The questionnaire was composed of two questions:

1. Among the matters discussed in geography lessons, indicate 3 (for each class level) in which students usually have learning and assimilation difficulties;
2. If you could use the computer as an aid in cartography studies, which subject matter in the classroom would be the most appropriate.

After the questionnaire application, a list of matters referred by teachers was made, which was based on the programmatic contents approved by the Ministry of Education of Guinea-Bissau to the discipline of social sciences. The categorization considered the following branches of geography: physical and human geography, cartography and environment. This first step allowed the selection of the topics that were part of the teaching modules, and also the definition of the school levels

---

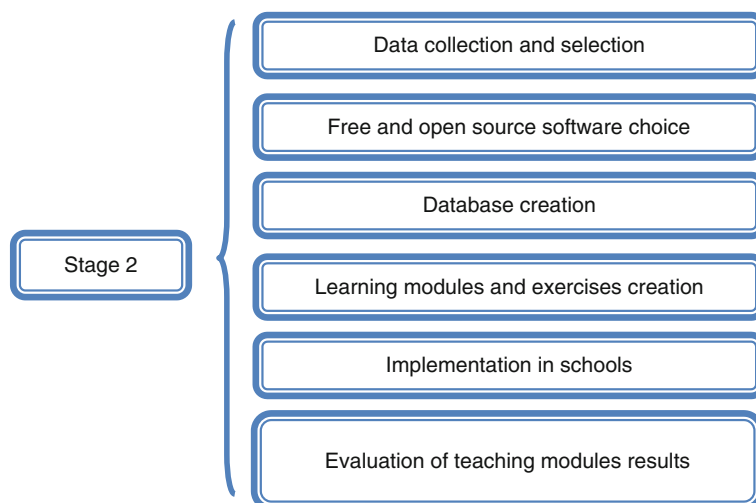[1] Courtesy of GeoEye Company Foundation for the project.

Fig. 2 Activities developed in the second stage of the methodology

in which the project would be implemented. The second stage of methodology followed as shown in Fig. 2.

a. Data collection and selection

This step was done in partnership with 5 geography teachers that participated in the questionnaire applied, in order to determine the best way to address the topics chosen.

b. Free and open source software choice

The goal of this step was to select a software to construct the databases. It should offer Portuguese interface, be of simple manipulation and accept various file formats. Terra View free GIS was chosen.

c. Database creation

Working with students' reality aroused more interest in performing the tasks, the GIS TerraView allowed them to visualize spatial phenomena and to find explanations for the subjects under study. For example, students could discuss and analyze the influence that the proximity to Bissau (the capital of the country) may have on other regions.

Due to outdated maps used in the geography classes, a database was done using recent documents from UNDP (UN Development Program), UNESCO (United Nations Educational, Scientific and Cultural Organization) and INE (National Statistics Institute of Guinea-Bissau). However, due to the lack of a cartographic base for the processing of such information in the GIS, we chose to work with countries that are part of the Economic and Monetary Union of West Africa
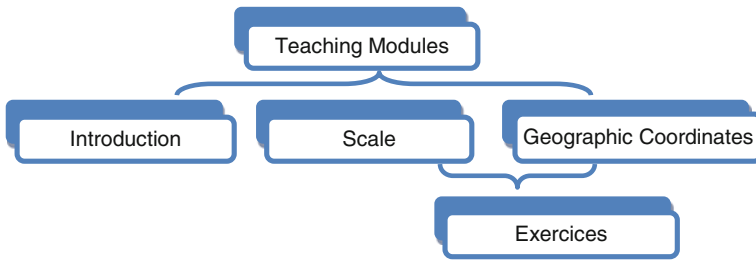
Fig. 3 Interface scheme of teaching modules

(UEMOA), which includes Guinea-Bissau. The following informations were selected to be part of the GIS database: gross domestic product, land area, population density and social indicators such as HDI, literacy rate, life expectancy and health indicators.

The base map from the database provided by ESRI was also used. The selected participating countries of the UEMOA were exported to TerraView GIS.

d. Learning module and exercises creation

After the definition of the school levels in which the project would be implemented and the content being explored, the modules, covering the topics geographic coordinates and scale were created, as shown in Fig. 3.

An introduction to define concepts such as: What is a map? What is it for? How is it built? What kind of data can be find on maps? - was included in the modules.

The teaching modules, called EACG (Teaching and Learning through Cartography and Geotechnology), were developed in flash application, so that it functioned as an electronic book, involving both theoretical and self executable exercises with solutions, as shown in the examples of Fig. 4.

At the end of each module (scale and geographic coordinates), the student was evaluated by a list of exercises of evolutionary character, ie, the difficulty increased as the student could find satisfactory answers in a previous question. The exercises were programmed so that the student could learn from his own mistake, as shown in Fig. 5.

f. Implementation in schools

A group of teachers evaluated the use of the teaching modules as supplementary material for geography lessons. In addition, students also evaluated the material through a questionnaire, based on theoretical material used and exercises from the teaching modules. Figure 6 shows the stages taken during the implementation of the project in the schools.

The *complementary activities* held along with the flash modules were made using Terra View GIS. The students used the database, made maps and observed their territory. The virtual globe Google Earth was also used in the *complementary activities*.
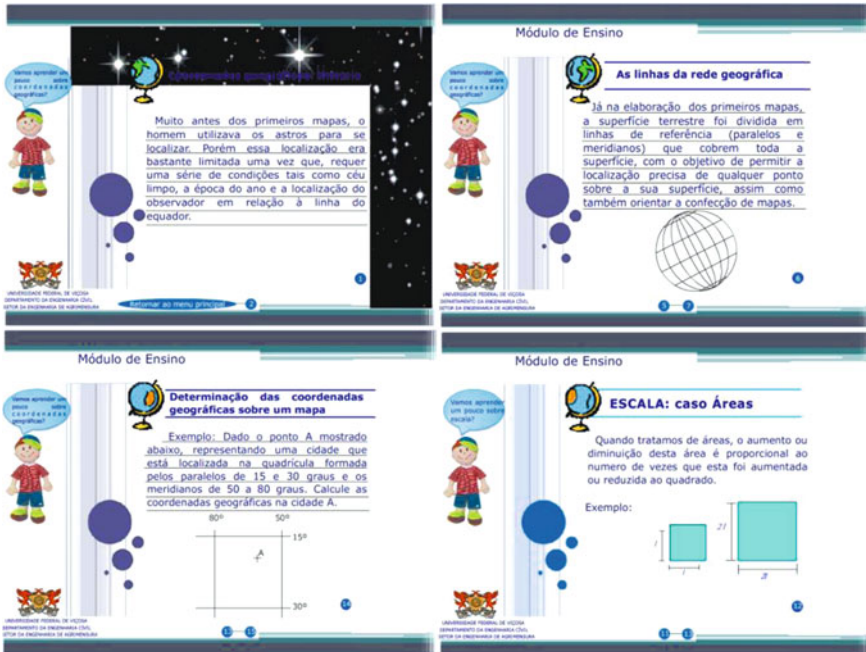
Fig. 4 Examples of EACG contents and activities

g. Evaluation of the results of teaching modules

To assess the impact of the modules on learning in the classes that participated in the study, the evaluation process was divided into two parts, the first was the application of a pretest and after the teaching modules application, students did a posttest.

The pretest was applied to assess the knowledge the students had on relevant matters to the module. The posttest was used to measure the real impact that the proposed modules had in the students' learning. Both tests were composed of six questions of multiple choices, other than those existing on the module.

After analyzing the results, the suggestions from students and teachers to improve the module and exercises were considered, so that the materials developed could meet their needs for technological resources for teaching purposes.

3 Results

The project was presented on the first contact with students and the pretest was applied (Fig. 7). One of the difficulties found during the implementation of this
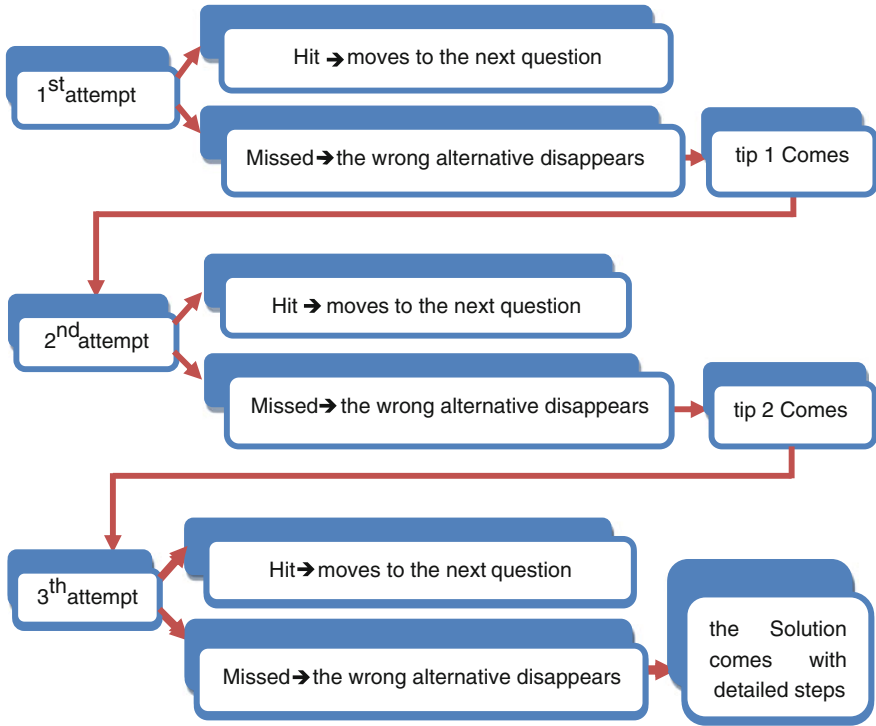
Fig. 5 Sequence of programmed exercises in the modules

first evaluation was the format of the test (multiple choice questions), since they were accustomed to discursive tests.

After the pretest application, the correction was made along with the students.

There were problems in the use of computer room because the schools do not yet have their own laboratory. Therefore, it was necessary to request permission from PASEG Organization (Support Program for Education of Guinea-Bissau), present in the main public high schools in Guinea-Bissau, to use its computer lab. The PASEG lab has 10 computers at Agostinho Neto lyceum and 9 computers at Kwame N'Krumah lyceum, where computer classes are offered to students and teachers.

It took 2 weeks to complete the modules due to schedules available in the PASEG computer Lab (Fig. 8).

In the *complementary activities*, the students produced thematic maps with the use of TerraView GIS and Google Earth.

(a) *In Google Earth*: the students showed interest in knowing the Guinean territory; many of them had no idea of the size of the regions in the country. Students also visited the schools and virtually saw their cartographic
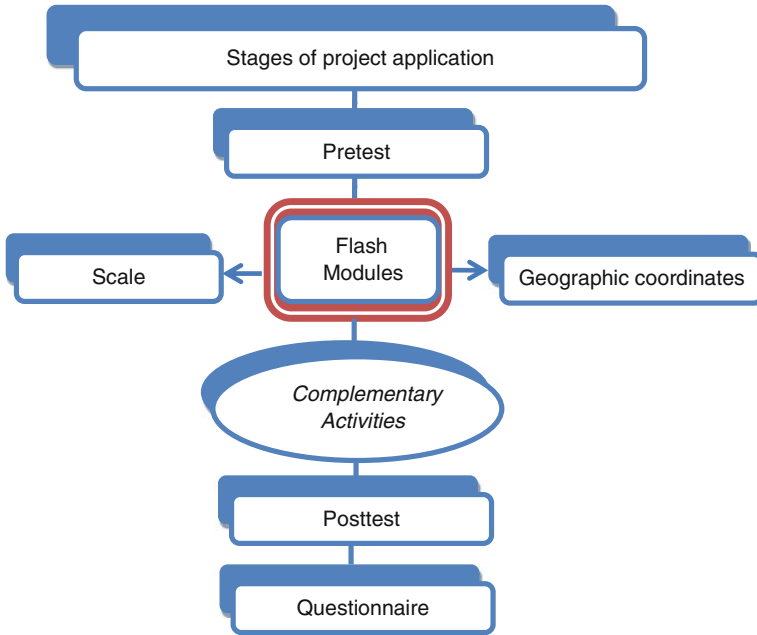
Fig. 6 Stages of the project application in Guinea-Bissau

Fig. 7 Pretest at the Institute Tchico Té and João XXIII Lyceum

representations. Moreover, they compared the city of Bissau with the city of Dakar and could perceive differences such as tree planting, paving of streets, the size of the city and rivers' identification.

(b) *In TerraView GIS*: first a short introduction was made on the program, where students could view the world map in different projections, for example. They prepared Guinea-Bissau's administrative regions maps and UEMOA thematic maps as shown in Fig. 9.

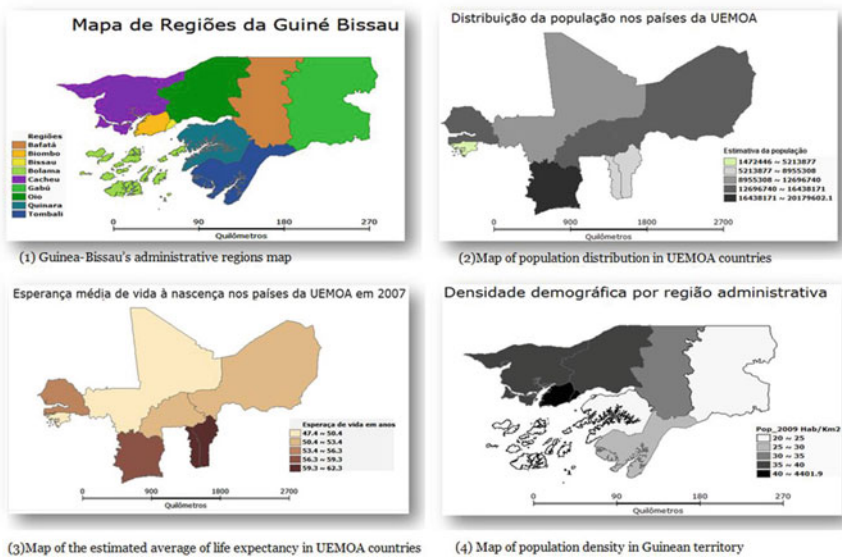**Fig. 8** Application of the modules in the 9th classes in Agostinho Neto Lyceum

Fig. 9 Maps produced by the students in Terra view GIS

3.1 Evaluation

A general analysis of the results showed that from the students of public lyceums that participated in both, the pretest and posttest (a total of 57 students), 29 successfully performed over 50 % of the posttest questions, compared with 5 in the pretest. That is, after the application of the teaching modules and activities using geotechnology, approximately 42 % of students who had previously done only half or less of the pretest's questions, managed to increase their performance, as shown in Table 2. More details about the statistics evaluations taken in all the classes can be found in Nosoline (2011).

It was possible to observe that:

Table 2 General analysis of results obtained in pre and post testing

Correct answers	Number of students	
	Pretest	Posttest
50 % or less	52	28
More than 50 %	5	29
Average right answers	1.39	3.42
Standard deviation	1.35	1.60

– The average increased 2.5 times compared to that obtained in the pretest;
– Forty-seven students obtained higher score in the posttest than in the pretest;
– Three students achieved the same score in pre-test and posttest;

In João XXIII lyceum, of the 63 participating students, 31 answered correctly more than 50 % of the posttest's questions. It was observed that the percentage of students who succeeded in up to half of the questions decreased approximately 24 %. In this lyceum the teaching modules were not purposely used in two classes in order to compare results.

It should be emphasized that in both samples shown, only the students who participated in both tests were considered in the comparison.

3.2 Qualitative Evaluation

Most students considered that the use of learning modules and geotechnology facilitated the topics' comprehension and aroused greater interest to learn due to a more active way of dealing with the issues presented. They felt more stimulated to do the tasks proposed.

This fact could be confirmed, since most students returned to the computer labs on their own initiative to remake the modules, even having already completed the activities, with the intention of obtaining a better performance in relation to that already achieved.

The teachers expressed satisfaction with the achieved results, praised the initiative and realized the importance of the use of geotechnology in schools as auxiliary tools in class, as reported by one of the teachers, when asked about the lessons taught:

> I consider these classes of extreme importance due to the adequacy of the methodology applied, which makes the content closer to reality.

In relation to the material developed for the application of the methodology, the same teacher stated that:

> The auxiliary material greatly helped in consolidating the concepts, as well as allowing to manipulate the computer; the schemes outlined here allowed a clear view and understanding of the issues covered.

And as a suggestion for the project improvement, the teachers indicated the production of the material on CD-ROM, to be used as reference material.

It is possible to infer that the use of geotechnology facilitated the understanding of the subjects studied for the sample used. Although, we cannot conclude that the performance improvement of these students resulted from the use of EACG teaching modules and classes taught with the support of Google Earth and TerraView. There could have been other factors that influenced the performance of students, as interest in computer and by innovations in class.

3.3 Difficulties Faced

The following describes some difficulties in the course of applying the methodology in the Schools.

1. The language

Portuguese is the official language of the country, but the difficulty with Portuguese language is still a problem existing in the routine of Guinean students, especially in public schools. This fact was confirmed at first contact with the students, so it was necessary to rethink communication, mixing the two languages, Portuguese and Creole,[2] in order to optimize the time available for classes.

The dynamics adopted for the evaluation was to apply two tests (pretest and posttest), teaching modules, complementary activities (GIS and virtual globe) and a questionnaire. In the two tests, especially in the pretest, students demonstrated great difficulty in understanding what was asked, so the Creole was chosen to end the doubts. When asked to respond to the questionnaire, this problem became more evident, as many students assigned grades not consistent enough with the written opinion.

2. Mathematics

Most students did not sympathize with mathematics itself, this explains why in the subsequent[3] school levels, the majority chooses groups from the human and biological areas. And despite this difficulty in all classes in public schools, when using examples involving money, all without exception, had success in the exercises with a logical and very fast correct answer, without the need to perform calculations on notebook or use the calculator.

---

[2] The association of Creole as a language is due to the fact that this is the integration dialect of all ethnicities existing in Guinea-Bissau, the most used in the communication process.

[3] From the 10th class students choose a group where the main subjects are listed according to the desired university course.

Thus, an adjustment was done in the way the questions would be resolved, a different approach was introduced in which the necessary calculations were related to money.

3. The relationship with students

Most of the students demonstrated timidity in responding or express an opinion on any matter. There are several factors that contributed to this, among them it can be highlighted the lack of familiarity with the researcher and the language of communication, because many are ashamed of making mistakes, whether in conjugation of verbs, or gender agreement, since, in Creole, there is not article to distinguish gender.

As a solution, the use of Creole was allowed, so the students could express their opinions more easily. It was also reported that the assessments would not be part of the official bimonthly evaluation of geography discipline.

4. Computers

Few students had already used the computer before that time. Therefore, it was necessary, in the first contact, make a brief introduction to computers use. The Project also contributed to the process of digital inclusion, as the accomplishment of proposed tasks was the first opportunity to familiarizing students and some teachers with computer technology.

4 Conclusion

The use of geotechnology and computer technology in the school proved to be a facilitator in understanding the subjects studied, thus providing a significant gain in student achievement, mainly when referring to forms of spatial representation and manipulation of geographic information.

In countries where there are lack of teaching materials, digital and free data would be a solution, so that teachers could create their teaching resources and prepare more attractive and dynamic lessons, using updated data and computer tools for interactivity.

However, what is lacking in Guinea-Bissau are more projects in the educational area, as this pioneering work in the country, which was for many students their first contact with the computer. It is hoped that this research had contributed to awaken educators and leaders to the importance of the use of low cost geotechnology and computational tools for spatial thinking as a teaching resource in African schools.

New technologies, particularly geotechnology, meet a set of tools with broad potential applications, leaving the teacher to explore and tailor it according to their needs, including social purposes.

Maps enable spatial domain and facilitate the understanding of spatial phenomena, for this reason, cartography and geography learning encourage people to

better understand their fundamental role in society. According to Santos (2007), a citizen is one who knows the space in which he is inserted. Knowing not only the geometric space, but also its social, political and economical organization, it is possible to have a minimum knowledge to criticize it and thus have a voice in the political struggle to influence and change its structure.

For future studies it would be interesting to increase the duration of projects, to develop new activities, analyze, for a longer period of time, the behaviour of students during the lessons for the adequacy of material and so measure the best use of technological resources in teaching and learning process. It would also be important to develop exercises that instigate logical reasoning, perform intermediate tests for measuring the retention time of the acquired knowledge and develop a course for teachers in order to ensure their effective participation in the projects.

Regarding the material used, it requires the creation of a more friendly interface that allows the integration of teaching modules (scale and geographic coordinates) with the GIS.

There were many difficulties to be faced, mainly because of the poor infrastructure of schools; these are the challenges for school systems, however, "*the underprivileged will experience the pleasure of breaking the isolation and enjoy their basic right to information, participation, growth*"(Maranhão 2002).

References

Bonham-Carter GF (1994) Geographic Information Sistem for geocientists: modeling with GIS. Elsevier, Kindlington, p 398

Hodgkiss AG (1981) Understanding maps: a systematic history of their use and development. Dawson, Folkestone, p 209

Katuta AM (2002) A linguagem cartográfica no ensino superior e básico. In: Oliveira AU, Pontuschka N (orgs) Geografia em perspectiva. São Paulo, Contexto, pp 133–139

Maranhão MA (2002) Telecomunidade e Exclusão Digital. Jornal Gazeta do Povo. Curitiba, PR. http://ufpa.br/imprensa/clipping%2005.02.2002.htm. Accessed 2 May 2002

Neves RJ, Cruz CBM (2007) O uso de representações gráficas geradas a partir de ferramentas de geoprocessamento nos estudos em sala de aula. Revista Brasileira de Cartografia No 59/01

Nosoline IM (2011) O uso das geotecnologias como recurso didático nas aulas de geografia. Master thesis, Universidade Federal de Viçosa. Viçosa

Santos M (2007) O Espaço do Cidadão. São Paulo, Brasil: Edusp

TERRAVIEW 4.0 (2010) São José dos Campos, SP: INPE. www.dpi.inpe.br/terraview. Accessed 2 June 2010