



OGRS

2012

OPEN SOURCE GEOSPATIAL
RESEARCH & EDUCATION
SYMPOSIUM

Symposium proceedings

Open Source Geospatial Research & Education Symposium

24th - 26th October 2012

**Yverdon-les-Bains
Switzerland**

www.ogrs2012.org

Edited by Olivier Ertz, Stéphane Joost
& Marj Tonini



SYMPOSIUM PROCEEDINGS

Open Source Geospatial Research & Education Symposium

Edited by Olivier Ertz, Stéphane Joost & Marj Tonini

Cover design and layout : Sandrine Divorne

Find this book on [lulu.com](https://www.lulu.com)



<http://creativecommons.org/licenses/by/3.0/>

These texts are free :

- to Share - to copy, distribute and transmit the work
- to Remix - to adapt the work
- to make commercial use of the work

Attribution : You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Waiver : Any of the above conditions can be waived if you get permission from the copyright holder.

Public Domain : Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.

Other Rights : In no way are any of the following rights affected by the license:

- Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations
- The author's moral rights
- Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights

Acknowledgements	10 - 11
About editors	13
Introduction	14 - 15

Organizers & Partners *19 - 21*

Program *25 - 31*

Committees *35 - 37*

Keynote talks *40 - 71*

Research conferences *75 - 290*

Education	75 - 101
Earth Science & Landscape	102 - 129
Data	130 - 168
Remote Sensing & Spatial Analysis	170 - 207
Urban Simulation	208 - 221
Tools	222 - 237
Posters	240 - 290

Workshops *294 - 320*

Discussion groups *324 - 327*

The second edition of the OGRS symposium would not have been possible without the contribution and involvement of the members of the organizing committee and without the support of their institutions: Prof Daniel Rappo from HEIG-VD/COMEM, Prof Florent Joerin from HEIG-VD/EC+G, Dr Devis Tuia from EPFL/LASIG, Dr Hy Dao from UNIGE/InfoGéo, Dr Erwan Bocher, Mr Gwendal Petit from CNRS/IRSTV and the editors of this book, Dr Marj Tonini from UNIL/CRET, Dr Stephane Joost from EPFL/LASIG and Prof Olivier Ertz from HEIG-VD/COMEM. Special thanks to Dr Grégory Giuliani from UNIGE/enviroSPACE for his important contribution in the preparation of the workshops. Also, many thanks to Prof Francis Grin (HEIG-VD/EC+G) and Prof François Golay (EPFL/LASIG) for their confidence and support.

OGRS 2012 is grateful to the SNSF (Swiss National Science Foundation), the Office for Economic Affairs (SPECo) and the ICT Competencies Network (RCSO-TIC) and the scientific committee of the University of Applied Sciences Western Switzerland (HES-SO) who facilitated the optimal organization of this event. We also want to express a strong word of appreciation to Dr Pierre Dessemontet and the city of Yverdon-les-Bains for ensuring a pleasant stay for attendees travelling from all over the world. A special thanks to Ms Sandrine Divorne, Ms Séverine Bise, Ms Anna Rita Bartolotta and all the HEIG-VD staff involved in the organization.

The organizing and program committee is most grateful to the members of the scientific advisory board for their active collaboration in the peer review phase. A special thanks to Prof Elizabeth Wentz for her help. Every scientific contribution was reviewed by two or three reviewers and revised accordingly to ensure quality. We would like to commend all the authors for their high quality work, which is the backbone of a great program. In particular, we would like to thank our four keynote speakers, Prof Helena Mitasova, Dr Gérard Hégron, Prof Sergio J. Rey and Prof Robert «Roebi» Weibel, for their wholehearted support of the OGRS adventure from the very start and for their scientific contributions which helped to the success of OGRS 2012.

Finally, many thanks to Mrs Catherine Hirsch and Dr Roland Prélaz-Droux, the Director and Deputy Director of the HEIG-VD, the Vaud School of Business and Engineering, for their help and support in hosting this symposium.

Olivier Ertz
Stéphane Joost
Marj Tonini

October 2012

Olivier Ertz is professor at the University Of Applied Sciences in Yverdon-les-Bains, Western Switzerland (HEIG-VD). His research and teaching activities are attached to the Media departement (COMEM) and concerns computer science and geospatial information technologies. His research contributes to the development of geospatial technologies to push GIS and cartography on the web (GeoWeb). His recent interests are focusing on : authoring and sharing of cartographic content in a collaborative context, portrayal interoperability through standards, polypublication of cartographic content in a multi-target and multi-channel context, cartographic content management systems and cartographic dashboarding.

Stéphane Joost is a quantitative geographer. He works as a Research and Teaching Associate at the Laboratory of Geographic Information Systems (LASIG), in the School of Architecture, Civil and Environmental Engineering (ENAC) of the Ecole Polytechnique Fédérale de Lausanne (EPFL). He is specialized in the contribution of Geographic Information Science for the conservation of plant and animal genetic resources, and for the understanding of the genetic mechanisms of evolution. He applies spatial analysis, spatial statistics, spatial modelling, geovisualization, and geocomputation methods to molecular ecology mainly (landscape genetics).

Marj Tonini is responsible of research at the Faculty of Geosciences and Environment of Lausanne University. She gained her PhD in 2002 at Sant'Anna School of advanced studies (Pisa, Italy), defending a thesis on agro-environmental modeling. Since 2005 she is working at Lausanne University where she collaborates with the Geomatics group as specialist in GIS and in space-time cluster analyses. She is in charge of courses in geomatics (GIS, 3D modeling). The research projects that she is currently carrying out concern: point process analyses and cluster detection of environmental phenomena (forest fires, landslides); wild urban interface detection.

Do you remember the Open Source Geospatial Research and Education Symposium (OGRS) in Nantes? "*Les Machines de l'Île*", the Big Elephant, the "*Storm Boat*" with Claramunt, Petit et al. (2009), and "*le Biniou et la Bombardé*"? A second edition of OGRS was promised, and that promise is now fulfilled in OGRS 2012, Yverdon-les-Bains, Switzerland, October 24-26, 2012. OGRS is a meeting dedicated to sharing knowledge, new solutions, methods, practices, ideas and trends in the field of geospatial information through the development and the use of free and open source software in both research and education.

In recent years, the development of geospatial free and open source software (GFOSS) has breathed new life into the geospatial domain. GFOSS has been extensively promoted by FOSS4G events, which evolved from meetings which gathered together interested GFOSS development communities to a standard business conference. More in line with the academic side of the FOSS4G conferences, OGRS is a rather neutral forum whose goal is to assemble a community whose main concern is to find new solutions by sharing knowledge and methods free of software license limits. This is why OGRS is primarily concerned with the academic world, though it also involves public institutions, organizations and companies interested in geospatial innovation. This symposium is therefore not an exhibition for presenting existing industrial software solutions, but an event we hope will act as a catalyst for research and innovation and new collaborations between research teams, public agencies and industries.

An educational aspect has recently been added to the content of the symposium. This important addition examines the knowledge triangle - research, education, and innovation - through the lens of how open source methods can improve education efficiency. Based on their experience, OGRS contributors bring to the table ideas on how open source training is likely to offer pedagogical advantages to equip students with the skills and knowledge necessary to succeed in tomorrow's geospatial labor market.

OGRS brings together a large collection of current innovative research projects from around the world, with the goal of examining how research uses

and contributes to open source initiatives. By presenting their research, OGRS contributors shed light on how the open-source approach impacts research, and vice-versa.

The organizers of the symposium wish to demonstrate how the use and development of open source software strengthen education, research and innovation in geospatial fields. To support this approach, the present proceedings propose thirty short papers grouped under the following thematic headings: Education, Earth Science & Landscape, Data, Remote Sensing, Spatial Analysis, Urban Simulation and Tools. These papers are preceded by the contributions of the four keynote speakers: Prof Helena Mitasova, Dr Gérard Hégron, Prof Sergio Rey and Prof Robert Weibel, who share their expertise in research and education in order to highlight the decisive advantages of openness over the limits imposed by the closed-source license system.

Here's wishing you an enriching reading!

Olivier Ertz
Stéphane Joost
Marj Tonini

ORGANIZERS & PARTNERS

heig-vd

Haute Ecole d'Ingénierie et de Gestion
du Canton de Vaud

www.heig-vd.ch

Hes·SO

Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

www.hes-so.ch

School of Business and Engineering Vaud (HEIG-VD)

With its 2000 students, the HEIG-VD is the largest school partner of the HES-SO (University of Applied Sciences, Western Switzerland). The campus of the HEIG-VD is based in Yverdon-les-Bains and is perfectly integrated into the economic environment.

Scientific studies with practical applications

At the HEIG-VD, theoretical knowledge is continuously linked with its implementation. The teaching provided by professors mainly from the economic sector and industry, encourages students to adopt a practical approach. They progress gradually to engineer or economist status, ready to create, conceive, innovate, control, frame or even to found their own company.

A recognized proximity with the economic environment

In the spirit of University of Applied Sciences policy, the HEIG-VD considers the transfer of knowledge to the economic environment as particularly important. The HEIG-VD Institutes are actively involved in the economy and industry through new product development, innovative solutions and management practices. The students are associated in this research through their practical diploma projects, whose topics are mostly proposed by companies. Once this training is acquired, it is possible to approach professional life with solid arguments.



UNIL | Université de Lausanne

UNIL - University of Lausanne

www.unil.ch



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

EPFL - Ecole Polytechnique
Fédérale de Lausanne

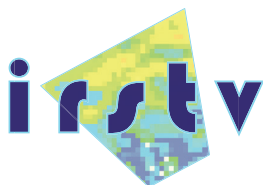
www.epfl.ch



UNIVERSITÉ
DE GENÈVE

University of Geneva

www.unige.ch



Institut de recherche en sciences et
techniques de la ville - FR CNRS 2488

IRSTV - Institute for Research
on Urban Sciences and
Techniques

www.irstv.fr



FNS / SNF - Swiss National
Science Foundation

www.snf.ch



SPECo - the Office for
Economic Affairs

www.vaud.ch



City of Yverdon-les-Bains

www.yverdon-les-bains.ch

PROGRAM

8:00	<p style="text-align: center;">Registration</p> <p style="text-align: center;">St-Roch building</p>		
9:00	<p style="text-align: center;">Room S153</p>	<p style="text-align: center;">Dynamic space-time visualisation: an introduction to i2maps</p> <p style="text-align: center;">Christian Kaiser</p>	<p style="text-align: center;">Room S155</p>
		<p style="text-align: center;">TinyOWS, the high performance WFS-T server</p> <p style="text-align: center;">Olivier Courtin, Vincent Picavet</p>	<p style="text-align: center;">Room S157</p>
			<p style="text-align: center;">Introduction to PostGIS: data management and geoprocessing</p> <p style="text-align: center;">Gregory Giulliani, Yaniss Guigoz</p>
12:30	<p style="text-align: center;">Lunch</p>		
14:00	<p style="text-align: center;">Room S153</p>	<p style="text-align: center;">Exploratory Spatial Data Analysis with PYSAL</p> <p style="text-align: center;">Sergio Rey</p>	<p style="text-align: center;">Room S155</p>
		<p style="text-align: center;">Exploit Pleiades PHR data with the ORFEO Toolbox library</p> <p style="text-align: center;">Manuel Grizonnet, Julien Michel</p>	<p style="text-align: center;">Room S157</p>
			<p style="text-align: center;">OrbisGIS : from GIS to standard-based geoservices</p> <p style="text-align: center;">Antoine Gourlay, Alexis Guéganno, Gwendall Petit, Erwan Bocher, Olivier Ertz</p>
17:30	<p style="text-align: center;">Ice breaker</p>		
18:30			

8:00	Registration – Cheseaux building	
9:00	Opening plenary – Aula <i>Chair: Pr. Olivier Ertz</i> Introduction and presentation Welcome message from Yverdon-les-Bains city council Pierre Dessemontet	
9:30	Building open source geospatial education at research universities: where we are and what is holding us back Pr. Helena Mitasova	
10:15	Coffee break	
10:45	Education – Aula	Earth Science & Landscape – Room F01
	A Field report on the role of Free and Open Source Geospatial Software at University of Applied Sciences Hans-Jörg Stark	r.rortstab: a GRASS-based deterministic model for deep-seated landslide susceptibility analysis over large areas Martin Mergili, Ivan Marchesini, Mauro Rossi, Fausto Guzzetti, Wolfgang Fellin
11:10	Hydroweb: an Open Source educational WebGIS platform for the understanding of spatio-temporal variations of meteorological parameters at the watershed scale Stéphane Joost, Jens Ingensand, Michael Kalbermatten, Rolf Peter Tanner, Vincent Luyet	Cluster analysis of geological point processes with R free software Marj Tonini, Antonio Abellán, Andrea Pedrazzini
11:35	Establishing a new Center of Excellence for the Free and Open Source Software for GIS (FOSS4G) Phillip Davis	Geographical analysis and numerical quantification of visual impact for aerogenerators and photovoltaic panels using Open Source GIS Annalisa Minelli, Ivan Marchesini, Pierluigi De Rosa, Luca Casagrande, Michele Cenci
12:00	Design and implementation of a distance education course on open source web mapping Barend Köbben, Ivana Ivanova	Relatedness and scale dependency in very high resolution digital elevation models derivatives Kevin Leempoel, Stéphane Joost
12:30	Lunch	
14:00		

14:00

Aula

The open source GIS, an ideal framework for the development of and integrated modelling platform devoted to sustainable urban planning: first steps with OrbisGIS and CartoPolis

Dr. Gérard Hégron

14:45

Poster session - Aula

Free Software and Open Source in Education: Geoinformatics at the CTU in Pragues

Martin Landa

Spatial patterns analysis of environmental data using R

Carmen Vega, Jean Golay, Marj Tonini, Mikhail Kanevski

Design and Creating a Two-way Public Participation Geographical Information System Platform: An Open Source Architecture

Antonio Jose Fernandes Da Silva, Jorge Gustavo Rocha

OSM indoor: moving forward

Jorge Rocha, Nair Alves

Multimodal Planner: From Prototype to Production

Francisco José Peñarrubia

The ESA BEAM Toolbox and Development Platform

Norman Fomferra, Daniel Odermatt, Carsten Brockmann, Peter Regner

WPS tools to support geological and geomorphological mapping

Ivan Marchesini, Mauro Rossi, Massimiliano Alvioli, Michele Santangelo, Mauro Cardinali, Paola Reichenbach, Francesca Ardizzone, Federica Fiorucci, Vinicio Balducci, Alessandro Cesare Mondini, Fausto Guzzetti

The Challenge of Geospatial Big Data Analysis

Teerayut Horanont, Apichon Witayangkurn, Shibusaki Ryosuke

Collaborative authoring and polypublication of cartographic content

Olivier Ertz, Julien Le Glaunec, Erwan Bocher

15:30

Coffee break

16:00

16:00

Discussion groups (all in parallel)**Urban remote sensing and environmental indicators***Chair: Pr. Elizabeth Wentz***Academic research and open source software to serve open standards development***Chair: Pr. Olivier Ertz***What licence should be used to release software or data produced in an academic context ?***Chair: Eric Grosso***Spatial Data Infrastructures in developing countries: Is Open Source a solution?***Chair: Dr. Gregory Giuliani***PGIS, from crowdsourcing to decision-making : research challenges, perspectives and emergences***Chair: Pr. Stéphane Roche***What about the future of OGRS symposium***Chair: Dr. Erwan Bacher*

17:00

Social event: «Bath time» at Yverdon Thermal Spa !*(free entrance included with the social event option)*

17:30

Social event: fine dining at the Grand Hôtel des Bains*(share a unique moment with your colleagues)*

20:00

8:30	Registration - Cheseaux building	
9:00	Aula	
9:50	Open source spatial analysis: lessons for research and education from PYSAL Pr. Sergio J. Rey	
10:15	Data - Aula Open source software for Big Data : Experiences in indexing and browsing geo-archival records Jefferson Robert Heard, Richard Marciano	Remote Sensing & Spatial Analysis - Room F01 An open tool to register landscape oblique images and generate their synthetic model Timothee Produit, Devis Tuia
10:40	Physical Landscape of Britain and Northern Ireland: technical development Claudio Piccinini, Mike Smith, Janet Hooke, Katherine Hesketh	Open tools and methods for large scale segmentation of Very High Resolution satellite images Julien Michel, Manuel Grizonnet, Arnaud Jaen, Sébastien Harasse, Luc Hermitte, Jonathan Guinet, Julien Malik, Mickaël Savinaud
11:10	Coffee break	
11:35	Using GRASS and PostgreSQL/PostGIS for the development of the automatic preprocessing for a distributed vector-based hydrological model Sonja Jankowsky, Pedro Sanzana, Flora Branger, Isabelle Braud, Yan Paillé, Florent Brossard, Fabrice Rodriguez	Utilization of the Scythe C++ open source library for statistical geocomputation in livestock landscape genomics Sylvie Stucki, Saif Agha, Menghua Li, Stéphane Joost
12:00	An open and powerful GIS data discovery engine Martin Ouellet, Stéfano Biondo	MCDA-GIS integration: an application in GRASS GIS 6.4 Gianluca Massej, Lucia Rocchi, Luisa Paolotti, Salvatore Greco, Antonio Boggia
12:30	Ontology Based Domain Specific Search of Crowdsourced OpenStreetMap Dataset and Wiki Stefan Keller, Michel Ott	How to make R, PostGIS and QGIS cooperate for statistical modelling duties: a case study on hedonic regressions Olivier Bonin
14:00	Lunch	

14:00	Aula	
	How open source can help achieving sustainability of e-learning content: The GITTA experience Pr. Robert Weibel	
14:45	Urban Simulation - Aula	Tools - Room F01
	Toward a coupling between GIS and agent simulation. USM : an OrbisGIS extension to model urban evolution at a large scale Frederic Rousseaux, Erwan Bocher, Antoine Gourlay, Gwendall Petit	GeoPeople project: using RESTful Web API to diffuse geohistorical database as open data Eric Grosso, Christine Plumejeaud, Benjamin Parent
15:10	Using open-source tools for the simulation of urban transportation systems Quoc Tuan Nguyen, Alain Bouju, Pascal Estraillier	An opensource tool to build urban noise maps in a GIS Nicolas Fortin, Erwan Bocher, Judicaël Picaut, Gwendall Petit, Guillaume Dutilleux
15:35	Coffee break	
16:00	Aula	
	Chair: Dr. Stéphane Joost	
	Feedbacks from discussion groups Joërin Florent	
16:45	Q/A, conclusions and further informations	
17:15	One last drink together with an extraordinary visit at Maison d'Ailleurs (all attendees are invited)	

COMMITTEES

PROGRAM & ORGANIZING COMMITTEE

(*) scientific committee co-chairs

- **Erwan Bocher***, IRSTV - FR CNRS 2488, *France*
- **Hy Dao**, University of Geneva, Dept. of Geography, *Switzerland*
- **Olivier Ertz***, HEIG-VD, University of Applied Sciences, Vaud, *Western Switzerland*
- **Gregory Giuliani***, University of Geneva, Institute for Environmental Sciences, *Switzerland*
- **François Golay**, EPFL, Geographic Information Systems Laboratory, *Switzerland*
- **Francis Grin**, HEIG-VD, University of Applied Sciences, Vaud, *Western Switzerland*
- **Florent Joerin**, HEIG-VD, University of Applied Sciences, Vaud, *Western Switzerland*
- **Stéphane Joost***, EPFL, Geographic Information Systems Laboratory, *Switzerland*
- **Gwendall Petit**, IRSTV - FR CNRS 2488, *France*
- **Daniel Rappo**, HEIG-VD, University of Applied Sciences, Vaud, *Western Switzerland*
- **Marj Tonini***, University of Lausanne, Institute of Geomatics and Risk Analysis, *Switzerland*
- **Devis Tuia**, EPFL, Geographic Information Systems Laboratory, *Switzerland*

SCIENTIFIC ADVISORY BOARD

- **Giuseppe Amatulli**, State University, Brookings, *South Dakota*
- **Bastian Baranski**, Institute for Geoinformatics, University of Münster, *Germany*
- **Thierry Badard**, Department of geomatic sciences, Laval University, *Canada*
- **Jean-Daniel Bonjour**, ENAC-IT, EPFL, *Switzerland*
- **Flora Branger**, CEMAGREF, *France*
- **Stefano Casalegno**, Stefano Casalegno, Environment and Sustainability Institute, University of Exeter, *UK*
- **Christophe Claramunt**, Naval Academy Research Institute, *France*
- **Arie Croitoru**, Dept. of Earth and Atmospheric Sciences, University of Alberta, *Canada*
- **Pierluigi deRosa**, Dept. of civil and environmental engineering, University of Perugia, *Italy*
- **Sara Fabrikant**, Department of Geography, University of Zurich, *Switzerland*
- **Jérôme Gensel**, Pierre Mendès France University, *France*
- **Gilles Gesquières**, LSIS laboratory, University of Provence, *France*
- **Carlos Granell**, European Commission Joint Research Centre, *Italy*
- **Manuel Grizonnet**, CNES, *France*
- **Eric Grosso**, COGIT Laboratory, IGN, *France*
- **Frédéric Hubert**, Department of geomatic sciences, Laval University, *Canada*
- **Ionut Iosifescu Enescu**, ETH Zürich, Institute of Cartography and Geoinformation, *Switzerland*

- **Ari Jolma**, Department of Civil and Environmental Engineering, Aalto University Lahti Center, *Finland*
- **Christian Kaiser**, Department of Geography, University of Zurich, *Switzerland*
- **Stefan Keller**, HSR University of Applied Sciences, Eastern *Switzerland*
- **Nathalie Long**, LIENS laboratory, La Rochelle University, *France*
- **Ivan Marchesini**, Geomorphology Research Group, CNR IRPI, *Italy*
- **Markus Neteler**, Fondazione Edmund Mach, Research and Innovation Centre, *Italy*
- **Judicaël Picaut**, IFSTTAR Institute, *France*
- **Nicolas Ray**, Enviroospace laboratory, Geneva University, *Switzerland*
- **Eduardo Corbelle Rico**, University of Santiago de Compostela, *Spain*
- **Stéphane Roche**, Department of geomatic sciences, Laval University, *Canada*
- **Frédéric Rousseaux**, LIENS laboratory, La Rochelle University, *France*
- **Christian Sallaberry**, University of Pau & Pays de l'Adour, *France*
- **Hans-Jorg Stark**, FHNW University of Applied Sciences and Arts, *Northwestern Switzerland*
- **Stefan Steiniger**, University of Calgary, Alberta, *Canada*
- **Elizabeth Wentz**, School Of Geographical Sciences and Urban Planning, *Arizona State University*
- **Li Xiang**, Department of Geography, East China Normal University, *China*
- **F.Javier Zarazaga-Soria**, University of Zaragoza, *Spain*

KEYNOTE TALKS



HELENA MITSOVA

Associate Professor since 2008 at Department of Marine, Earth and Atmospheric Sciences (MEAS), North Carolina State University (NCSSU), Raleigh.

Her teaching and research areas are about spatial Modeling and visualization : modeling and monitoring of landscape processes, evolution of coastal topography, geoinformation science, multi-temporal lidar data processing, spatial interpolation and topographic analysis, coastal and watershed erosion studies. Applications of GIS and multidimensional dynamic cartography for sustainable land use management and conservation of natural resources. Member of the OSGeo foundation and Open Source GRASS GIS development team.



GÉRARD HÉGRON

French representative and scientifique expert at UERA - Urban Europe Research Alliance. Scientific Director since 2008 in charge of sustainable city at IFSTTAR (French Institute of Science and Technology for Transport, Planning and Networks).

From 2000 to 2011 he was the director of IRSTV (Research Institute on Urban Sciences and Techniques) gathering 20 laboratories working on interdisciplinary research projects about urban environment and sustainable urban planning and where spatial information and open source GIS development (OrbisGIS) plays a major role for spatial analysis and for the integration of urban data and models. He is also currently the director of the national scientific network on Urban Modeling.



ROBERT WEIBEL

Professor of Geographical Information Science since 2000 at Department of Geography, University of Zurich.

His research interests concern, computational cartography, spatio-temporal analysis, spatial analysis for the cultural sciences and GIS for the environment. Also, he has been a principal investigator of the GITTA project, Geographic Information Technology Training Alliance, a platform offering e-learning content as Open Educational Resources. Initiated in 2001, the GITTA project won in 2008 the Medida Prix for innovative use of digital media in education.



SERGIO REY

Sergio Rey (Ph.D., University of California, Santa Barbara) is a Professor in the School of Geographical Sciences and Urban Planning at Arizona State University, where he also serves as an executive committee member and core research faculty in the GeoDa Center for Geospatial Analysis and Computation.

Rey's research interests focus on the development, implementation, and application of advanced methods of spatial and space-time data analysis in the social sciences. His substantive foci include regional inequality, convergence and growth dynamics as well as neighborhood change, segregation dynamics, spatial criminology and industrial networks. Rey is the creator and lead developer of the open source package STARS: Space-Time Analysis of Regional Systems as well as PySAL: A Python Library for Spatial Analysis. In 2010 he co-edited with Luc Anselin, *Perspectives on Spatial Data Analysis* (Berlin: Springer).

Building open source geospatial education at research universities: where we are and what is holding us back

Authors

- **Helena Mitasova**, Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC, *USA*
- **Martin Landa**, Department of Mapping and Cartography, Faculty of Civil Engineering, Czech Technical University, Prague, *Czech Republic*
- **Makiko Shukunobe**, Center for Earth Observation, North Carolina State University, Raleigh, NC, *USA*

KEYWORDS : geoinformatics program, course material, graduate certificate, GRASS GIS, PostGIS

Introduction

In spite of a growing government and industry support for open source software only few university Geospatial Information Science curricula include open source approach, and most programs focus on use of proprietary software. At the same time, many science and engineering disciplines require flexibility and portability provided by the Free and Open Source Software (FOSS) to support coupling of complex computational tools with geospatial data processing. Some of the largest technology companies, such as IBM and Google, contribute to FOSS development and many U.S. government agencies now require open source as a condition for grants and contracts. FOSS also provides a viable alternative for agencies in regions with limited resources. Moreno-Sanchez [1], in his editorial for special issue of Transactions in GIS devoted to FOSS4G 2011 conference, highlights FOSS4G as a mature alternative in the field of geospatial technologies.

Overview of existing efforts

FOSS4G is used in academia in a variety of settings. The most widespread but hard to track are applications and development in university research laboratories. Related work has been presented annually at FOSS4G conference academic sessions and selected papers were published in special issues of refereed scientific journals [1], [2], [3], [4]. Short courses, workshops, and summer schools is another area with numerous successful efforts and extensive material available through ELOGeo [5] and OSGeo Edu [6] inventories and on the annually updated OSGeo Live DVD. Special topics courses focused on FOSS4G are commonly offered as electives at research universities and colleges all over the world, but there are only few continuously updated, established courses [7]. Academic communities most active in FOSS4G are still geographically concentrated in Europe (especially in Italy and Spain), North America, and Japan although there are increasing efforts to support adaptation of FOSS/FOSS4G in other regions, especially in the developing world (FOSS4G 2008 in South Africa is a good example).

FOSS4G at North Carolina State University (NCSU)

NCSU is located in a region with thriving FOSS industry lead by RedHat, the first FOSS company to reach 1 billion in annual revenues in 2011, validating the viability of the FOSS business model. NCSU has a rapidly growing Geospatial Information Science and Technology (GIST) program. It includes Graduate Certificate, Professional Masters of Science in GIST, and several PhD programs that incorporate GIST-focused research. Major expansion of the program into Geospatial Analytics is under way. The program focus is on general GIST concepts and hands-on practice in both open source and proprietary systems. Advanced technology for in-class courses and for Internet-based delivery is used to support and enhance the instructor work and student experience, both in the classroom and through distance education. The GIST program is interdisciplinary (there is no Geography or Geomatics department at NCSU), managed by the Center for Earth Observation at College of Natural Resources, with faculty coming from 20 Departments and 8 Colleges. The program covers FOSS4G in one core geospatial analysis course that includes GRASS GIS and a number of elective courses that include python, PostGIS and WebGIS applications from the OpenGeo software stack. Easy to install FOSS4G binaries were key for introduction of open source into the curriculum.

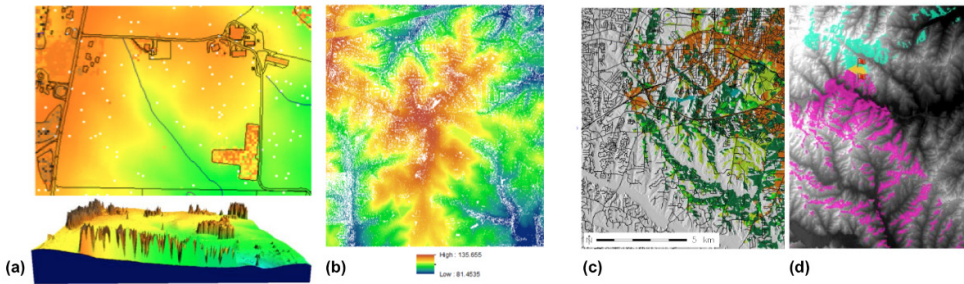


FIGURE 1

Assignment on lidar data processing. Binning point cloud to raster in (a) GRASS and (b) ArcGIS. GRASS output includes cross-section of bare ground and first return surfaces. Analyzing viewsheds: (c) land use composition in the visible area with GRASS, (d) visibility between two points displayed in ArcGIS

To illustrate the integrated approach, specific examples from the graduate course on Geospatial Modeling and Analysis are provided. In this course, the lectures, focused on concepts and methods, are software independent. The related assignments are simultaneously performed in GRASS GIS and ArcGIS, with slight differences in tasks to make the results more interesting, as illustrated by several examples (Figure 1). The course material, with design © by NCSU, and content under CC license, is available free online and updated annually to reflect changes in new software releases. Takehome midterm exam allows students to chose software to accomplish the tasks, leading to interesting and often creative results (Figure 2).

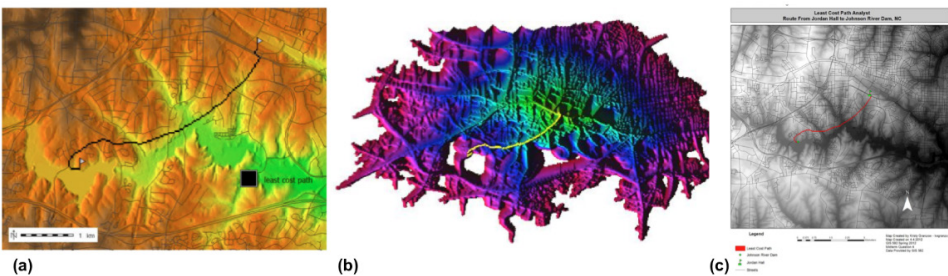


FIGURE 2

Midterm exam: results of the least cost path task performed and visualized in: (a) GRASS GIS with path displayed over shaded relief, (b) GRASS GIS with path draped over the 3D representation of cost surface, (c) ArcGIS displayed over elevation raster

Important component of the course is an independent project: again students chose the software and it is quite common that both ArcGIS and GRASS are used. Several examples of students' projects that use FOSS4G demonstrate the diverse open source geospatial applications (Figure 3), such as creating a DEM by using the mobile phone GPS, analysis of karst topography from lidar data, or study of coastal dynamics and hurricane impacts. Several projects lead to peer reviewed publications [8], [9]. Although the students come into class with at least some experience in ArcGIS, GRASS GIS and open source concept itself have been new for all and many students were eager to gain the unique knowledge and skills that FOSS4G offers. As additional benefit, the students working as research assistants introduced FOSS4G into their research laboratories.

Under the auspices of the Memorandum of Understanding between the

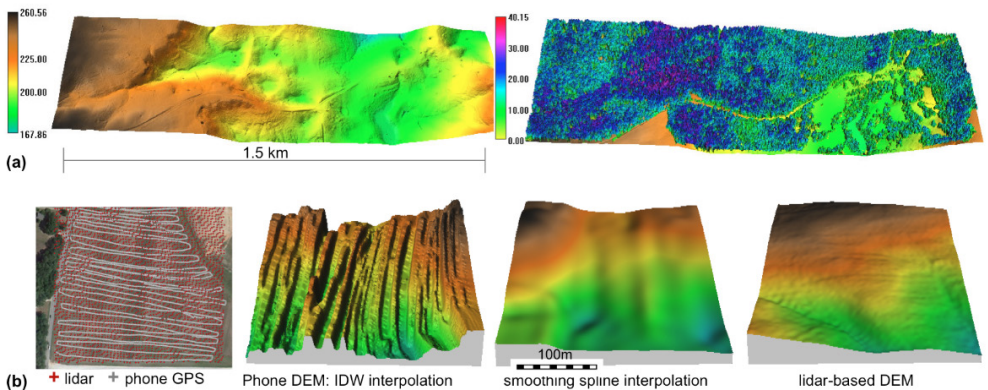


FIGURE 3

Geospatial Modeling and Analysis course projects using GRASS GIS: (a) Using LiDAR to map sinkholes and canopy height in Mammoth Cave National Park; (b) creating a digital elevation model using mobile phone GPS

International Cartographic Association and OSGeo [10], there are plans to establish OSGeo Research and Education Laboratory (OSGeo REL) at NCSU with the goal “to support the development of open-source geospatial software technologies, training and expertise and to provide support for building up the development of open-source GIS teaching and training materials in order

to encourage wider participation globally.” The active collaboration with the Geoinformatics program at the Czech Technical University and its planned OSGeo REL will explore effective approaches to international collaboration to achieve these goals that could then be applied on broader scale.

FOSS4G at Czech Technical University (CTU) in Prague

FOSS4G tools have been used in courses at the Faculty of Civil Engineering, Czech Technical University (CTU) in Prague since 2005. First, the assignments for a Remote Sensing course in the Geoinformatics program were developed with GRASS GIS. A new course called Free Software GIS was introduced later, in 2007. Within this course, the students learn about the FOSS4G environment in general, including a community-driven development, OSGeo role, as well as the desktop and webbased projects. The course is available also in English to make it accessible for international Erasmus students. Thanks to Professor Cepek, FOSS tools have been used in the Geoinformatics and Geodesy and Cartography programs for several years, including PostgreSQL database system in the Introduction into Database Systems course, Qt framework in Programming in C++ course and Introduction into GNU/Linux OS course has also been offered. The students can further enhance their knowledge in database systems in the Introduction into Geospatial Processing course focused on geospatial data handling in database systems using PostGIS. Thanks to these efforts the students get valuable experience with various FOSS4G tools (programming, geospatial database systems, GIS desktop or web-based applications) in addition to the proprietary GIS solutions. Several students became involved in FOSS as translators, power users and developers. Since 2008 several Geoinformatics students from CTU participated in Google Summer of Code (GSoC) program and developed a new GUI for GRASS visualization module wxNviz (in 2008, 2010, and 2011, [Figure 4](#)) and for a network analysis module wxVNet in 2012. Other students have been involved in QGIS development as their semester projects or diploma thesis (e.g., recently developed QGIS plugin for Czech cadastral data or Workflow builder for QGIS Sextante project). These activities and the annual Geoinformatics conference along with GRASS GIS community sprint held in Prague provide a sound basis for the planned establishment of a regional OSGeo Research and Education Laboratory at CTU.

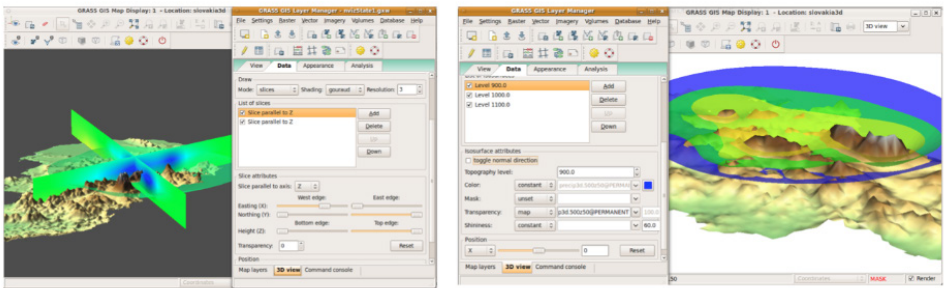


FIGURE 4

GUI for 3D visualization developed as GSoC project

Conclusion and future directions

In spite of several successful programs [7], incorporation of FOSS4G into GIST education remains challenging. Although substantial educational material is available, preparing and maintaining free, up to date, on-line educational material for multiple systems requires substantial resources in time and expertise but the number of faculty directly involved in FOSS4G projects is still rather small. Student contributions to FOSS4G remain limited in spite of positive role of GSoC and university research laboratories. Several initiatives based on experiences from universities with successful programs [7] may bring progress on these issues. The proposed global network of OSGEO Research and Education Laboratories [10], if supported by sufficient funding, has a potential to create a coordinated front for FOSS4G education material development, including electronic textbooks, task oriented software tutorials, lectures and webinars. Direct involvement of university faculty and students in the FOSS4G projects as power users or developers and broader participation in GSoC as mentors and programmers could provide the bridge between the academic world and FOSS4G development community. Testing of new modules, development of add-on tools and contribution to teaching materials can be incorporated into courses as class assignments, helping students to overcome potential barriers to involvement in FOSS4G communities.

As the successful programs demonstrate, FOSS4G provides a cost-effective, portable and highly functional toolset for interdisciplinary research and education that is especially valuable in science and engineering programs. With coordinated effort and collaboration of academic community it can be fully integrated into GIST curricula worldwide and bring a new level of innovation and creativity into geoinformatics and geomatics education.

Acknowledgment

We would like to thank the students Alet Terblanche, Ali Durmaz, Kristy Granzow, and Anna Kratochvilova for the samples of their class and project work used in the figures.

[1] MORENO-SANCHEZ, R. Free and Open Source Software for Geospatial Applications: A Mature Alternative in the Geospatial Technologies Arena. *Transactions in GIS* 16 (2012), 81–88.

[2] NETELER M., AND MITASOVA H., GRASS as Open Source - Free Software GIS: accomplishments and perspectives. *Transactions in GIS* 8(2) 2004, 145-154

[3] COETZEE, S; COOPER, A K.; FLEMING, G; NETTERBERG, I, Free and Open Source Software for Geospatial (FOSS4G), *Special issue of South African Computer Journal* 43, 2009, 1-2.

[4] BROVELLI, M, MITASOVA, H, NETELER, M, RAGHAVAN, V, Free and open source desktop and Web GIS solutions, *Applied Geomatics* 4(2), 2012, 65-66

[5] E-LEARNING FOR THE OPEN GEOSPATIAL COMMUNITY. ELOGeo Repository. *The University of Nottingham website*. Retrieved Sept 19, 2012 from <http://elogeio.nottingham.ac.uk/xmlui/>

[6] OSGeo Educational Content Inventory. *Open Source Geospatial Foundation website*. Retrieved Sept 19, 2012 from http://www.osgeo.org/educational_content

[7] MITASOVA, H. University courses based on Open Source Geospatial Software. *NCSU Marine Earth and Atmospheric Sciences website*. Retrieved Sept 19, 2012 from http://skagit.meas.ncsu.edu/~helena/grasswork/osgeoedu_courses.html

- [8] ALVAREZ J., MITASOVA H., ALLEN H.L., Estimating monthly solar radiation in South-Central Chile, *Chilean Journal of Agricultural Research* 71(4), 2011, 601-609.
- [9] CLINCH, A.S., RUSS, E., OLIVER R.C., MITASOVA H., AND OVERTON, M.F., Remote sensing estimation of Hurricane Irene soundside surge elevations on the North Carolina Outer Banks, *Shore and Beach* 80(2), 2012.
- [10] GARTNER, G, AND CHRISTL, A. 2011. Memorandum of Understanding between the Open Source Geospatial Foundation and International Cartographic Association. *Open Source Geospatial Foundation Wiki*. Retrieved Sept 19, 2012 from http://wiki.osgeo.org/wiki/MOU_ICA

How open source can help achieving sustainability of e-learning content: The GITTA experience

Authors

- **Robert Weibel**, University of Zurich, Department of Geography, *Switzerland*
- **Susanne Bleisch**, University of Applied Sciences Northwestern Switzerland (FHNW), Department of Infrastructure Engineering, University of Melbourne, *Australia*
- **Joël Fisler**, University of Zurich, Computing Services, *Switzerland*

Introduction

In the early 2000s, government initiatives were implemented in various countries that funded the development of e-learning content for academic programs. Common to these funding programs was the objective to promote the use of new media and information technology in teaching and learning in higher education, and the creation of synergies (and thus reduced costs) through reuse of teaching materials. Thus, typically, projects funded under these programs were required to consist of several networked partner groups. After funding had run out in the mid-2000s, however, many of the above projects quietly disappeared, or are no longer maintained, and exorbitant expectations raised in the hype over e learning in the early years had to be corrected. It is interesting to note that at the same time, there are several programs that have been existing for many years and continue to be successful. These are, however, degree-granting programs subject to tuition and not freely available.

Our hypothesis is that the failure of the government-funded projects is mainly due to two factors. First, there was typically a lack of integration into a curricular context (while the tuition-based projects are all firmly embedded

in Masters degree programs). And second, many government-funded projects were developed with the limited timeframe of the funding period in mind, and thus not built for sustainability and easy maintainability.

In this paper, we focus on the second issue, and particularly on how open source and content strategies may help in improving the longevity of e-learning projects. We will use the example of GITTA (Geographic Information Technology Training Alliance), a large, multilingual e-learning project. Among others, GITTA has spawned off eLML (eLesson Markup Language), an XML framework initially developed within GITTA, redesigned through the use in further e-learning projects and released as open source.

An overview of GITTA

GITTA is a multi-lingual project originally developed by a consortium of ten groups in several GIScience related disciplines at seven partner universities in Switzerland, and maintained by the GITTA Association since 2006. In the early years since its start in 2001, the project was funded through the Swiss Virtual Campus, a program of the Swiss federal government. The project's mission is to develop and deliver e-learning content for integration into higher education curricula in GIScience, preferably in blended learning mode [5]. In 2008, the project won the MedidaPrix, the largest award for new educational media in Europe.

Content: GITTA covers a large scope of topics in GI science and technology. Over 40 lessons exist today, grouped into six broad thematic modules: GI Systems, Data Capture, Data Management, Spatial Modeling, Spatial Analysis, and Data Presentation. Lessons focus on conveying theoretical and technical concepts and are organized into a basic and an intermediate level, depending on the level of complexity of concepts presented. Every lesson accounts for 2-3 hours of learning time for the student.

Besides the theory lessons, 6 case studies have been created in which students are confronted with realistic sample cases. In these case studies, the practical problem solving skills of the students are trained [4]. In each case study, a description of the assignment and original data material is provided. The student is then expected to develop solutions for solving the assignment using GIS tools.

Finally, the GITTA website also provides access to 15 lessons from the CartouCHE project (an e-learning project developed by a subgroup of the GITTA project partners), specializing on multimedia cartography and location-based services .

Pedagogical approach: The pedagogical design of GITTA is modular, but ensures at the same that a common structure is used for lessons and case studies, respectively [7]. Lessons and case studies are provided as a pool of learning resources rather than a closed, static course that follows a fixed order. These resources are provided as IMS or SCORM content packages, such that they can be imported to learning management systems (LMS).

The common pedagogical structure for theory-related content is shown in Figure 1. Levels (basic, intermediate) and thematic modules, as described above, provide the coarse grained structure that embeds lessons and helps structure their use in courses. Each lesson then consists of different units, which contain learning elements that follow the ECLASS model (Entry, Clarify, Look, Act, Self-assess, Summary), an extended version of the model originally developed by [2].

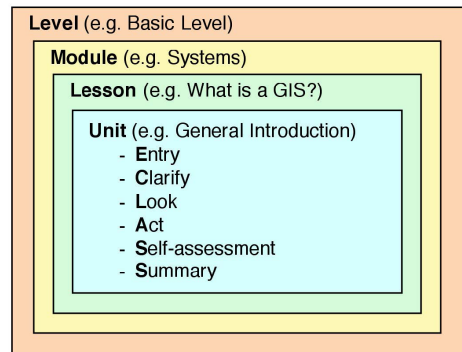


FIGURE 1

Organisation of GITTA content and pedagogical structure (ECLASS model)

Languages: Owing to the multilingual background of the original GITTA Consortium, content is available in different languages. On the basic level, lessons and case studies are available in German or French and partly in English. Lessons on the intermediate level are offered in English only. Translation of all non-English lessons to English is currently under way. Further languages (particularly French and Spanish) may follow, depending on funding. The technical infrastructure of GITTA has been designed such that it can support

different language versions, and to facilitate translation to additional languages (cf. next section).

Technical framework

eLML – the eLesson Markup Language: One of GITTA’s requirements was that the pedagogical model outlined above could be directly mapped to the technical infrastructure used for content development. However, when the project started in 2001 there were no formats, content standards or LMS available that could meet our needs for open standards, easy portability, and coherent structuring and presentation of content. It was then decided to base the content development on XML and subsequently, eLML was developed as a framework based on XML Schema. eLML was documented for general use and published as an open source project under the Apache License on Sourceforge.net since 2004. Since then a constantly growing number of projects and authors in Switzerland and other European countries started using eLML (e.g. a tutorial for InfoVis). Since 2006 eLML is maintained by the Computing Services of the University of Zurich and became an independent and successful open source project in the field of content creation, winning several awards. An in-depth description of eLML is given in [1]. Furthermore, the eLML website provides the complete technical documentation, downloads, and links to other projects using eLML. It also provides information about Firedocs, a WYSIWYG eLML editor implemented as Mozilla Firefox AddOn, to help content authors who are not versed in writing content in a markup language, as well as other productivity tools.

eLML and the ECLASS model [2] are intimately related. eLML implements ECLASS in XML by offering XML tags that follow the content and structure of ECLASS. Besides the known elements of the ECLASS model, eLML contains additional semantic elements such as glossary, bibliography, and metadata [1]. Hence, eLML offers all the necessary elements to develop e learning lessons such that 1) content follows a consistent form across the entire course, and 2) authors are provided with a guideline how lessons or case studies should be structured. Both conditions are absolutely crucial for content development and maintainability in large, distributed, multi-author projects such as GITTA.

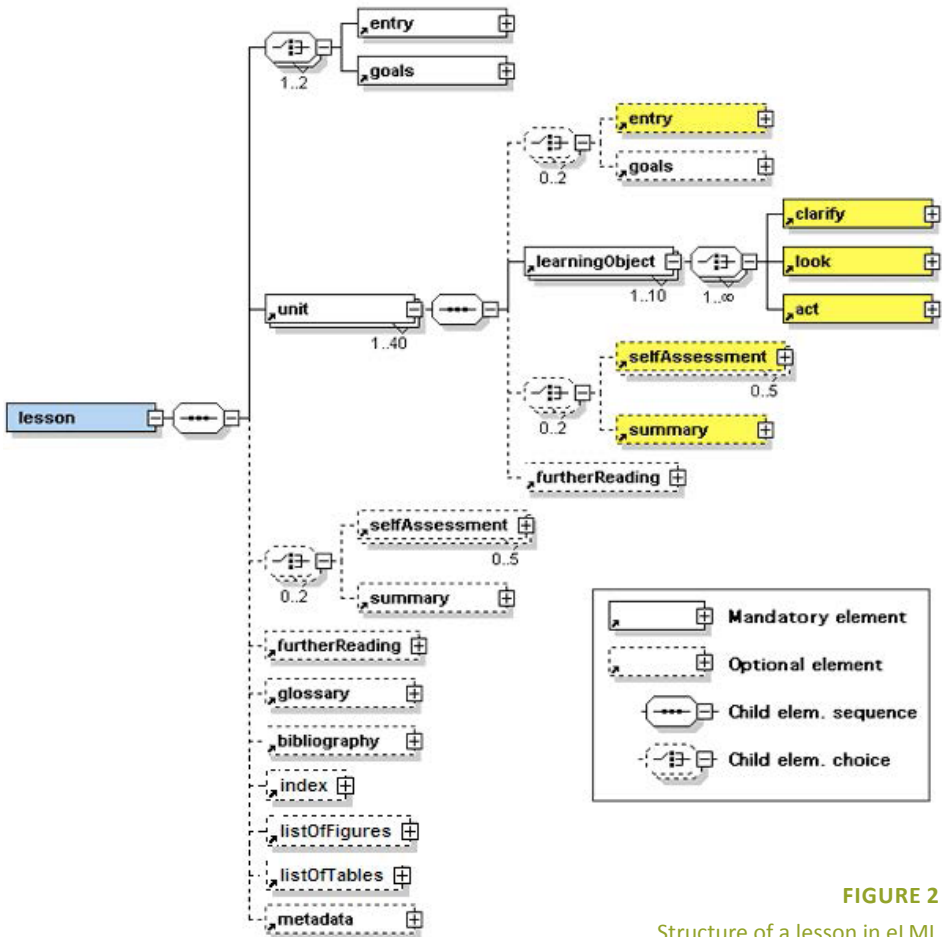


FIGURE 2

Structure of a lesson in eLML

Figure 2 shows the structure of an eLML lesson. The elements of the ECLASS model compose the structure of the eLML lesson and its units. The first structural layer of a lesson consists of an introduction (entry), the learning objectives of the lesson (goals), and the units contained in a lesson. Additionally, a lesson may contain a self-assessment section, a summary, a further reading section, glossary, bibliography, and metadata (providing descriptive information such as author, technical requirements, estimated time needed for completion, etc.).

The units of a lesson may contain an entry section, the goals, a self-assessment and a further reading section. They must contain at least one learning object (LO), composed of at least one clarify, look or act section. Additional content elements are available for micro structuring of content for presentation, such as markups for tables, lists, boxes or interactions like multimedia content or popups [1].

Lessons written in eLML (i.e. XML) must be transformed into a file format that makes it possible to display the content in commonly used web browsers or LMS. In this regard many options exist. It is also possible to create archiving formats such as IMS or SCORM content packages, for import into LMS such as Moodle, Blackboard, or OLAT. For mobile learning, mobile views of eLML lessons can be generated based on the jQuery Mobile framework. Hence, platform independence is truly achieved.

It is also possible to create archiving formats such as IMS or SCORM content packages, for import into LMS such as Moodle, Blackboard, or OLAT. Hence, platform independence is truly given.

Content management: As mentioned above, content in GITTA is not directly embedded into an LMS. Instead, it is held as a pool of lessons and case studies, each of which are offered in XML as well as derived formats. Hence, content is maintained on a content server that is running the open source Concurrent Versions System (CVS; [8]). The regular GITTA website offers a web front-end to the CVS content server, making it possible to access the latest version of individual lessons or case studies, either as XHTML or as content packages (IMS, SCORM) for integration in an LMS of choice.

Open content strategy

Originally, the GITTA consortium was planned as a closed organization, with the intention of sharing content only among the seven participating Swiss universities. Later on, the consortium agreed on adopting an open content strategy, releasing GITTA content under the 'by-nc-sa' Creative Commons license as of early 2006.

Two main reasons led to the adoption of this strategy. First, releasing materials as open educational resources (OER) made life a lot easier by simplifying legal matters between the consortium members as well as in communication with external users. Second, open content provided a major element of the strategy to achieve a long life for the materials created during the project phase. By publishing content as OER it was hoped to create a 'GITTA Community' that would use and maintain the released lessons, just like in other open source/content projects. A survey conducted among the GITTA subscribers in 2010 not only documented the worldwide usage of GITTA content but also provided valuable feedback. Among others, requests for translations to other languages were made, and new lessons or modifications of existing lessons proposed. More than forty respondents indicated that they would be willing to help with the further development of GITTA content, as translators, reviewers of existing lessons, authors of new or extended lessons, or developers of animations. This pool of potential volunteers represents a valuable resource, but managing these potential authors represents a major challenge, given the limited staff resources available in the member groups of the GITTA Association.

Conclusion

We have presented a summary of the open source and open content strategy of the e-learning project GITTA and its associated XML framework eLML. We are convinced that this approach has helped improving GITTA's sustainability, as it is one of the few non-commercial e-learning projects in GIScience that has continued to exist for many years. Besides perhaps ELOGeo, which however appears to be a relatively loose repository of resources with an educational bearing rather than a full-fledged e-learning project. The content of GITTA is routinely used and has received positive ratings from external sources (e.g. Medida Prix 2008 or [3]). The technical framework, in particular working with a platform and LMS independent approach based on eLML has facilitated the open content strategy and ensured sustainability. Otherwise, lessons developed in an LMS of 2001 would almost certainly no longer be accessible today. As a consequence of this good example, eLML has found use in various projects beyond GITTA. In an academic environment, OER and content sharing is certainly the way to go, even irrespective of sustainability considerations. The

GITTA content, information about the GITTA Association, and publications about the project can be found on the project website www.gitta.info. For more detailed information on the sustainability strategy the reader is referred to [6]. The educational use and curriculum integration of GITTA content is discussed in [7]. More information on eLML can be found in [1] and on the project website www.elml.org.

[1] FISLER, J., & BLEISCH, S. eLML, the eLesson Markup Language: Developing Sustainable e Learning Content Using an Open Source XML Framework. In WEBIST 2006 - International Conference on Web Information Systems and Technologies, Setubal, Portugal, April 11th-13th 2006. Retrieved Sept 9, 2012 from http://www.gitta.info/website/en/download/gitta/webist2006/WEBIST2006_eLML.pdf

[2] GERSON, S.M. E CLASS: Creating a Guide to Online Course Development For Distance Learning Faculty. Online Journal of Distance Learning Administration 3(4) (2000). Retrieved Sept 9, 2012 from <http://www.westga.edu/~distance/ojdla/winter34/gerson34.html>

[3] KATTERFELD, C. & KÖNIG, G. Analysis of E-learning Software and Guidelines for Quality Assurance in Photogrammetry, Remote Sensing and GIS. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Congress Beijing 2008, Vol. XXXVII, Part B6a, 45-53. Retrieved Sept 9, 2012 from http://isprsserv.ifp.uni-stuttgart.de/proceedings/XXXVII/congress/6a_pdf/2_WG-VI-2/02.pdf

[4] NIEDERHUBER, M., HEINIMANN, H. R. & HEBEL, B. e-Learning basierte Fallstudien zur akademischen Ausbildung in der Geoinformatik: Methodisches Konzept, Umsetzung und Erfahrungen. In J. M. Haake, U. Lucke & D. Tavangarian (Eds.): DeLFI 2005: 3. Deutsche e-Learning Fachtagung Informatik, Lecture Notes in Informatics 66 (2005), 34-45.

[5] THE ELEARNING GUILD. The eLearning Guild's Handbook of e Learning Strategy, 2007. Retrieved Sept 9, 2012 from <http://www.elearningguild.com/content.cfm?selection=doc.817>

[6] WEIBEL, R., BLEISCH, S., NEBIKER, S., FISLER, J., GROSSMANN, T., NIEDERHUBER, M. & HURNI, L. Achieving Sustainability in e-Learning for GIScience – Experiences from the GITTA Project. *Geomatica* 63(2) (2009), 109-118.

[7] WEIBEL, R., LÜSCHER, P., NIEDERHUBER, M., GROSSMANN, T. & BLEISCH, S. Delivering GIScience education via blended learning: The GITTA experience. In Unwin, D., Foote, K., Tate, N. & DiBiase, D. (eds.): *Teaching Geographic Information Science and Technology in Higher Education* (2012), Wiley-Blackwell, pp. 405-420.

[8] VERSPERMAN, J. *Essential CVS*, Second Edition. O'Reilly, 2006.

The open source GIS, an ideal framework for the development of an integrated modelling platform devoted to sustainable urban planning: first steps with OrbisGIS and CartoPolis

Authors

- **G rard H groun**, IFSTTAR/IRSTV, *France*
- **Erwan Bocher**, CNRS /IRSTV, *France*
- **Gwendall Petit**, CNRS/IRSTV, *France*

Sustainable urban planning : an interdisciplinary and systemic approach

Cities provide one of the major challenges to global sustainability. Over half the world's population is now urban and it is growing rapidly. Growth is driven by globalisation making a significant contribution to global environmental change. Cities affect social, economic, cultural and environmental sustainability. In this context, new and large-scale research is vital, to study not only cities, but also regions and their environments at all spatial scales, and over a range of timescales. Such research is challenging and complex, and will need an interdisciplinary approach. It needs to bring together the humanities and social sciences with the other major disciplines, and involve researchers, users and stakeholders in the process. It will embrace research that is quantitative and qualitative, and it may involve anything ranging from the setting up of urban observatories, to sophisticated modelling of urban systems.

In this context, sustainable urban planning needs the development of a systemic approach involving environmental, social and economic dimensions where the evaluation of public policies is of paramount importance for decision making. This approach requires setting up an integrated modelling platform for the

computation of indicators and spatial analysis tools at different spatial and time scales. A Spatial Data Infrastructure (SDI) integrating a Geographic Information System (GIS) enables the use of spatial data in an efficient and flexible way and remains an ideal framework to fulfill this objective.

This idea is emerging since the 2000s where GIS are more and more used for interdisciplinary and multisectoral contexts. The development of standards for data description and data exchange (interoperability) and the arrival of the concept of SDI, facilitate the interconnection of systems and the carrying out of systemic approaches [1][2]. Such integrated platforms can be developed by the actors of geographic information to manage the whole knowledge about a territory and to observe its dynamics (observatory). SDI are being multiplied at all levels of decision (European, national, regional, local) and in different institutional groups (public services, laboratories, etc.). In this way, at IRSTV (Institute for Research on Urban Sciences and Techniques) which is a research federation gathering 20 laboratories and leading an interdisciplinary research devoted to the development of knowledge, models, tools, methodologies for urban design and management, and to the evaluation of sustainable city policies, we have developed such an approach around an open source SDI, CartoPOLIS, and an open source GIS, OrbisGIS.

Urban data representation and management: need of an SDI

The understanding of natural and anthropogenic phenomena at work on a territory is essential to monitor, control and manage urban systems and their dynamics. Data are the key. Their massive acquisition and treatments go through the design and implementation of new methodologies and technologies adapted to different spatial and time scales. These new techniques should produce reliable data, making them easily accessible, allowing the crossing of various kinds of data to perform multi-criteria analysis. A Spatial Data Infrastructure (SDI) based on ISO and OGC standards, is suitable for the acquisition, processing, capitalization, sharing and preservation of urban spatial data and metadata. But to face systemic issues and interdisciplinary applications, several technical and scientific locks must be removed:

→ Description and storage of multi-source and multi-scale data and metadata

- Querying of such data: What kind of language for users?
- Utilization of such data by models or other tools: How to enable the integration?
- Provision of data and tools: Is the set of data standards sufficient to manage the data flows? How do we manage the data processing flows and the storage of the spatial analysis tools? How to visualize the spatial data according to specific mapping rules?

Cartopolis attempts to answer these questions [3]. This open source SDI consists of 5 modules:

- A single database aggregating all data collected or produced
- A server of data flow (Geoservice): this application provides geographic data via the Internet using the standard WMS (Web Map Services)
- A tool for cataloging data (Geocatalog): it consists of a set of metadata sheets structured according to ISO 19115. It contains information such as the temporal extent of the dataset, its spatial extent, its origin, its semantic features, etc
- A mapping Internet gate with graphical interfaces for querying the Geocatalog. The user can enter a word for retrieving a data or search for given data by specifying a bounding box or a temporal scope
- An open source Geographic Information System (GIS) called OrbisGIS [3]

Diagnosis and evaluation of sustainable urban policies: need of an open source integrated platform including a GIS

Sustainable urban planning relies on the understanding of complex systems and underlying phenomena. Modeling as a tool for understanding, diagnosing and forecasting becomes essential to the development and comparison of planning scenarios (evaluations) and to clarifying decision making. Therefore, we need a platform integrating a lot of urban models exploiting, sharing and producing spatial data (SDI).

A GIS linked to the appropriate SDI remains the ideal framework where data processing involves the cooperation between urban models (computation of indicators for instance) and where the spatial analysis of the results can be achieved. OrbisGIS has been defined and carried out to fulfil these goals [3].

The open source paradigm and the use of standards are essential to facing the integration of multidisciplinary packages and its use for very different professional applications:

- The exchange of data between urban models, applications, visualization tools, etc., requires interoperability and the use of standardized data. In our platform we use OGC standards (WMS, WFS, etc.) and ISO standards (19115, 19139, etc.)
- Integrating new models as plugins in a GIS must be facilitated and transferable from one GIS platform to another one. This necessitates a modular approach to development: the developer should opt for the model-view-controller software design pattern
- The language for data query and data processing must be comprehensible to different users and must be easily completed when adding additional features like new evaluation tools or models. For this reason, OrbisGIS integrates an extended spatial language (which includes the management of both vector and raster data) based on the Simple Feature SQL (SFS) standard [4][5]. This language offers a unique way to describe spatial processing. Coupled with the Web Processing Service it permits to share in a common platform all geospatial processing available on-demand via internet. This would enhance the current uses of the SDI by pooling all the processes built by IRSTV researchers and engineers (noise mapping, flooding modeling, atmospheric pollutant dispersion, etc.) and by creating a geospatial knowledge repository to study cities [6]
- The visualization of spatial data must cope with static, dynamic and multi-scale data, must be comprehensible by all stakeholders (good representation of the semantic and use of standards), and must be easily edited and disseminated on different supports. Symbology encoding standards have to be consistent with the purposes of research and new applications. IRSTV is

working in this direction at OGC in cooperation with the IICT laboratory [7]. As a proof of concept, a first evaluation of cartographic publications with the Geospatial PDF format has been performed in the framework of OrbisGIS [8]

Collaborative development of the open source platform

At IRSTV, our goal was to build a common platform where each member of the Institute would be an active contributor, a common tool adopted by all the different users, a single platform able to integrate and capitalize all the methods and tools necessary to describe, model and manage the city. The development of such a platform needs a specific collaborative methodology to manage the understanding and the interaction between the various stakeholders (developers, users) and to preserve the quality of code.

Co-operate software engineering

To develop a common open source platform a collaborative approach for the specification of new features and new tools, a project management for the integration of new packages and the preservation of the quality of the delivered code (validation process) are essential. To achieve these goals in our platform, the quality of developments in the core, but also additional developments made by the community in the form of plugins, are controlled by the open source tool Jenkins to ensure a continuous integration without regression. The application source code is managed using the platform GitHub and the GIT tool, a tool for the sharing and the distributed control of codes.

Pedagogy and student training: from software engineering to participating in coding

Besides the development of this platform, IRSTV is engaged in academic (MSc on Sciences and techniques for urban planning) and professional training providing full courses to learn how to use and develop these tools. In 2001, we introduced a semester of 128 hours based on the use of open source software (GIS, SDI, relational database ...) and on the use of information from open data, plus a 160 hour training project dedicated to the engineering of geographic information. These training periods are conducted in cooperation with other existing communities like GvSIG and OpenStreetMap.

A research platform for proof and benchmarking

For a researcher, especially in the sciences of geographic information, the development of one's research proposals in an open source platform is a complementary means to prove to the scientific community the quality of its implementation and its usage, its computation performances according to the amount of data, and its ability to adapt to other datasets. It is also a way to compare its performances to other approaches on the same datasets. Finally, it is a way to demonstrate its relevance for the professional community.

Conclusion

As demonstrated in this paper, the open source paradigm addresses perfectly the needs and development methodology of an integrated modelling platform dedicated to sustainable urban planning. Participants coming from different scientific fields need to share the same urban data: open data and open source SDIs are an obvious choice. To address the systemic issue of urbanism, collaborators must integrate their various modelling packages in a single platform. These packages should be able to interact via a common language. New functionalities should be implemented to cross-examine newly generated data, and common symbology encoding will help standardize the visualization and communication of results. Research on sustainable urban development is a really great opportunity for innovation and progress in the open source community.

[1] STEINIGER, S., HUNTER, A J.S. *Free and Open Source GIS Software for Building a Spatial Data Infrastructure. Geospatial Free and Open Source Software in the 21st Century Lectures Notes in Geoinformation and Cartography*, (LNG&C) series, Springer-Verlag, 2009.

[2] GSDI. 2009. Spatial Data Infrastructure Cookbook. Global Spatial Data Infrastructure Association. Retrived september 15, 2012 <http://www.gsdi.org/gsdicookbookindex>

[3] BOCHER, E., PETIT, G. *OrbisGIS : Geographical Information System designed by and for research, Innovative Software Development in GIS*, ISTE & Wiley, Geographical Information System series, 2012.

[4] BOCHER E., LEDUC T., MOREAU G., GONZALÉZ -CORTÉS F., GDMS: An abstraction layer to enhance Spatial Data Infrastructures usability. 11th AGILE International Conference on Geographic Information Science (Girone, Espagne, 2008).

[5] LEDUC, T., BOCHER, E., GONZALÉZ -CORTÉS, F., MOREAU, G. *GDMS-R: A mixed SQL to manage raster and vector data*, Proceedings of GIS 2009 (Ostrava, République Tchèque, 2009).

[6] GOURLAY, A., AND BOCHER, E. Towards a WPS platform dedicated to an urban knowledge infrastructure. INSPIRE Conference (Edinburgh, Scotland, 27th June to 1st July, 2011).

[7] BOCHER, E., ERTZ, O., LAURENT, M., PETIT, G., RAPPO D. 2011. Cartographie et standard: du modèle à l'utilisateur. International Cartographic Conference (ICC) (Paris, France, 2011). Retrieved septembre 18, 2012 http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/E1-Open%20source%20ans%20web%20services/CO-404.pdf

[8] TOINON, V. 2012. Évaluation de publications cartographiques au format Geospatial PDF, Retrieved september 18, 2012 <http://mappemonde.mgm.fr/index.html>

Open source spatial analysis: lessons for research and education from PySAL (Extended Abstract)

Author

Sergio J. Rey, GeoDa Center for Geospatial Analysis and Computation. School of Geographical Sciences and Urban Planning, Arizona State University, *USA*

KEYWORDS : Open source, spatial analysis, education, research

Abstract: This paper explores the intersection of open source with the areas of spatial analysis research and education. Drawing on lessons learned in the development of PySAL: Python Library for Spatial Analysis, it touches on the opportunities and challenges related to the adoption of open source practices and culture. While open source has had major impacts on pedagogy and research in spatial analysis, these are somewhat under-appreciated and at times seen as separate spheres. The paper reconsiders open source spatial analysis teaching and research from an integrated perspective and suggests some possible future developments.

1. Introduction

The open source revolution continues to have major impacts on science and education and the field of spatial analysis is no exception. A number of recent overviews of open source spatial analysis and geographic information science have recently appeared [3, 4, 6, 2, 7, 1] and my intent here is not to provide a similar comprehensive coverage of this area but rather to build on a particular set of themes originally raised by Rey [4]. I do so by drawing on the lessons learned in the development and evolution of the PySAL project [5] as it has intersected with my teaching and research activities.

2. Lessons for Education

It goes against the grain of modern education to teach students to program. What fun is there to making plans, acquiring discipline, organizing thoughts, devoting attention to detail, and learning to be self-critical. Alan Perlis

Open source software and practices can have major empowering impacts on pedagogy. The free availability of the software offers a number of advantages in lab based courses. No longer are the students constrained to working in the school laboratory as they can now use the software installed on their own personal laptops, or home desktops, to complete exercises. This also allows for a greater degree of exploration and discovery by the student working by themselves and at their own pace.

These represent potential pedagogical wins for open source in geospatial education. My personal recent experience is that we still have a ways to go before these benefits are fully realized. During the fall of 2011 in my introductory course in GIScience, I decided to use QGIS as the software for the lab component in place of our traditional package of ArcGIS. This was something I had contemplated doing for quite sometime, but I always held back as the feature set and polish of QGIS were not yet at the stage where I felt comfortable doing so. By fall 2011, this had changed as the development of QGIS had reached an impressive state.

To my surprise, this switch was less than well received by the students. Emblematic of the main complaint was the following comment I received on an anonymous teaching evaluation: *I took this course as I heard we would be taught ArcGIS. I don't care about the science and the algorithms underneath the software, I want a job when I leave this class.*

While there were a minority of students who told me they appreciated the introduction to an open source alternative, the vast majority of the students were not happy about the switch. In addition, I received push-back from some of my colleagues who were concerned that not covering ArcGIS threatened relationships with community internship partners that had been carefully cultivated over the years.

I was completely blindsided by these responses and felt a mixture of disappointment and puzzlement. In hindsight, I admit that these potential negative impacts never entered my decision making calculus. At the same time, while I now see that these are pressing concerns, they also raise some important questions regarding the role of geospatial education. On the one hand, the current demands in the labor market for students trained in ArcGIS reflects the reality that previous generations of students we have trained in this software are now in key positions in these agencies and companies. Additionally, many of these agencies have invested much time and resources in their GIS infrastructures and are understandably conservative regarding any changes. But, what about the future? Is our task to train students for today's labor market or to equip them with the skill sets and knowledge so that they are ready for, and can create, the future geospatial labor market?

A second general lesson for geospatial education concerns graduate education and the seemingly the ironic situation of an embarrassment of riches in terms of freely available high quality programming tools for geospatial research on the one hand, and on the other, a general lack of desire to do any programming. I believe this stems from the challenges facing geography graduate students in that they not only need to acquire knowledge of substantive and methodological areas of the discipline but also somehow become proficient in programming. We have done a fairly poor job on the latter with solutions running from recommending introductory courses in computer science departments to learning on the job as part of a research project. The former is rather inefficient as my experience is geography students taking most introductory computer science classes come away without any idea of how to apply core concepts to geographical problems. The mentoring approach scores higher on this point, but does not really scale well.

Taken these together I think that the reality of the situation of open source and geographic education is currently rather mixed. At the undergraduate level the impact has been much more limited than I would have originally believed, due mainly to the institutional factors raised above. The situation is more evolved at the graduate level. Here I've seen several instances where access to the source code in PySAL has enabled a motivated graduate student to gain a deeper

understanding of a particular spatial analytical method. In hindsight, this mixed success may also suggest that a certain level of training and education may be required before the benefits of open source software can be experienced by students.

3. Lessons for Research

Open source has also had major impacts on research in GIScience. In the US, this is

clearly seen in research proposals to federal agencies as increasingly there are requirements that publicly funded projects include data and results management components so that subsequent research projects can replicate and extend funded projects. Having served on review panels for some of these agencies, a clear trend is that open source has been relied upon by many scientists to respond to these requirements. It should be emphasised that open source software offers clear advantages when it comes to replication as there are no longer any “black boxes” that conceal the implementation of a particular method or algorithm [8].

In an open source world it is also more likely that the scientific questions lead the way forward and the software itself is enhanced or modified to address these questions. This is in contrast to the proprietary world where the core programs themselves are not malleable. In the past this has led to the choice of research question being constrained by the functionality provided by the software. While the addition of scripting languages and plug-in or toolbox architectures do offer mechanisms for adding new functionality to address new types of questions, the issue of the social returns must also be considered.

For the academic engaged in open source software development, there are a number of challenges that are encountered. A chief one regards the academic evaluation and promotion system which places heavy emphasis on scientific publications. Development, maintenance and documentation of an open source spatial analysis package requires a significant investment in one's time and this cuts into time that could go towards writing and submitting journal articles

and books. For a package that becomes widely adopted there is the possibility that scientists who use the package in their own research take care to cite the package, but my hunch is this is done less often than one would like. There have been positive developments in this regard with journals such as Journal of Statistical Software that provide an outlet dedicated to developments in statistical software. It also reflects a shift in attitudes towards scientific software in that it is seen as scholarly work that should come under peer-review.

4. Conclusion

You think you know when you can learn, are more sure when you can write, even more when you can teach, but certain when you can program. Alan Perlis

Although the lessons outlined above treated education and research separately, this was for purposes of exposition only. There are clearly strong synergies between education and research as the quote from Perlis suggests. At the same time, there are some challenges that can hinder these synergies. One of our overriding goals in the development of PySAL has been to keep the level of code readability as high as possible, and here we have relied on the clear syntax of the Python language. We have always felt that the code can serve as a powerful source of information for students interested to learn the exact manner in which a spatial analytical method was implemented. While we have by and large kept to this goal, we have encountered tensions along the way. Keeping the code readable has required that we limit the number of third party libraries that PySAL requires. These libraries are often written in lower level languages such as C, C++, or Fortran and can offer substantial speed gains over pure Python implementations. At the same time the lower level code can be more difficult for the newly initiated spatial analyst to decipher. Faced with this trade-off, we have chosen pedagogy over speed.

4.0.1 Acknowledgments

This research was partially supported by NSF Award OCI-1047917, SI2-SSI: CyberGIS Software Integration for Sustained Geospatial Innovation.

- [1] BROVELLI, M., MITASOVA, H., NETELER, M., AND RAGHAVAN, V. Free and open source desktop and Web GIS solutions. *Applied Geomatics* (2012), 1–2.
- [2] HU, Q., CHEN, Y., ZHOU, Y., AND XIONG, C. An overview on open source GIS software with its typical applications. *Geomatics World 1* (2009).
- [3] RAMSEY, P. The state of open source GIS. In *Free and Open Source Software for Geospatial (FOSS4G) conference* (2007).
- [4] REY, S. J. Show me the code: Spatial analysis and open source. *Journal of Geographical Systems 11* (2009), 191–207.
- [5] REY, S. J., AND ANSELIN, L. PySAL: A Python library of spatial analytical methods. In *Handbook of Applied Spatial Analysis*, M. M. Fischer and A. Getis, Eds. Springer, 2010, pp. 175–193.
- [6] STEINIGER, S., AND BOCHER, E. An overview on current free and open source desktop gis developments. *International Journal of Geographical Information Science 23*, 10 (2009), 1345–1370.
- [7] STEINIGER, S., AND HUNTER, A. Free and open source GIS software for building a spatial data infrastructure. In *Geospatial Free and Open Source Software in the 21st Century: Proceedings of the first Open Source Geospatial Research Symposium, 2009* (2012), E. Bocher and M. Neteler, Eds., Springer, pp. 247–261.
- [8] YALTA, A. T., AND YALTA, A. Y. Should economists use open source software for doing research? *Computational Economics 35* (2010), 371–394.

RESEARCH CONFERENCES

A Field Report on the Role of Free and Open Source Geospatial Software at the University of Applied Sciences

Author

Hans-Jörg Stark, University of Applied Sciences Northwestern Switzerland School of Architecture, Civil Engineering and Geomatics Institute of Geomatics Engineering, *Switzerland*

KEYWORDS : University level GIS education, free and open source geospatial software, secondary school

Summary

Universities of Applied Sciences in Switzerland serve a fourfold purpose. In addition to education, they work in the field of applied research and projects. During the last few years, free and open source geospatial software (FOSGS) has proved to be a catalyst not only in the implementation of projects but also in their acquisition process. Further, it has built a bridge from researcher to students and supported a better integration of students' work into research work and vice versa. This paper presents a field report based on a few projects which were implemented at the Institute of Geomatics Engineering at the University of Applied Sciences and Arts Northwestern Switzerland FHNW.

Introduction

At the Institute of Geomatics Engineering (IGE) at the University of Applied Sciences and Arts Northwestern Switzerland FHNW, free and open source geospatial software (FOSGS) has been used since 2005. Starting with UMN MapServer, PostgreSQL/PostGIS, PHP and JavaScript for web-based applications [1], the number of FOSGS has grown over the years. One major project which has its roots at IGE is OpenAddresses.org [2], which was originally developed as a Google Maps mash up but was later migrated in 2010 to MapFish [3], which is

a framework for building rich web-mapping applications [4]. Due to these public web-based applications which were realised at IGE, public interest in FOSGS grew in the community and networks which are linked to IGE. This led to the situation where IGE was specifically contacted with the focus to realise research projects and applications with FOSGS. As a result, IGE was able to set up two large projects with the Bau- und Umweltdirektion Baselland [5], [6]. A service to evaluate the conformity of OpenGeoSpatial webservices was conceptionally designed with the 'Coordination, Geo-Information and Services (COGIS)' department of swisstopo, the federal geo-information center of Switzerland and is currently in the process of being implemented. With all these projects, fundamental skills for using Django and Python as the programming languages were developed. Additionally, GeoExt and OpenLayers were two of the main frameworks which were used.

Bridging gaps with free and open source geospatial software

Our know-how of Django, Python and GeoExt was the basis on which the public web-based application 'See You' was built. See You is a project designed for secondary school children [7] and is a further development of the Map Your World project [8]. Secondary school pupils carry a GPS logger for a week or two, upload their data into the project's central database from where different geospatial information is provided: a heat map that shows colour-coded accumulated GPS tracks, hot spots which indicate locations with either high frequency or where people meet and finally indoor locations which indicate potential indoor leisure or residential places are computed. The GPS tracks can be filtered by age, gender, time of day and days of the week.

Many aspects concerning different technologies within this project are covered both for those who use the application but also for the project supplier. Using low-cost GPS devices leads to the problem of how to pre-process this data to improve its quality for further analysis in the See You project. This task was transferred into students' projects with the support of research assistants at IGE. Not only [9] did students in the bachelor and master programmes work conceptionally on the issue but they were also able to use Python to test their concept. Their Python code can now be revised and easily integrated into the

project's architecture which is based on Python. Other issues that were dealt with similarly were the creation of the heat or density maps based on intensive literature research ([10],[11],[12],[13]) and the extraction of indoor locations. Thanks to the practice of sharing code and samples which is one of the main pillars in open source software, the creation of the heat maps did not have to be started from scratch but used a script provided by Seth Golub [14].

Python is one of the languages which can be easily learned and thus helped, especially in the See You project, to bridge the gap between research and the bachelor and master programmes at IGE. For students it is very motivating to see their work is not an artificial task but is integrated into real-life projects and that their effort is a contribution to a larger context. Free and open source geospatial software strongly supports this approach because both institutions like universities and students are able to access software licenses and evaluate them at no external cost. The software and quite often also samples are available and members of newsgroups or mailing lists are helpful and quick to reply when there are questions and obstacles. The project See You also shows that applying FOSGS had an even further positive impact on education in secondary school. If the mentioned application had needed to be set up with proprietary software, it might have failed due to the incurred cost.

Conclusion

In the author's experience, FOSGS has proved to be a catalyst in successfully implementing projects in an academic environment. This is due not only to the free software licenses but also thanks to the availability, the openness and willingness to support of communities doing open source software projects. Academic research and teaching experience have shown FOSGS to be a fruitful and productive environment to work in.

In addition, the Institute of Geomatics Engineering receives generous support from proprietary software vendors. Thus the future is not going to be 'only FOSGS' but rather a 'both FOSGS and proprietary GIS software' approach. Most of IGE's students will face proprietary software in their future careers, especially in the area of surveying. Nevertheless, the FOSGS which they came into contact with during their studies have opened their minds to another,

mostly complementary and occasionally even better alternative to the still very famous and proprietary software programmes which partly dominates the market. For the future, we are confident we will integrate FOSGS more and more into the teaching at IGE and balance the supply of different geospatial software.

[1] STARK, HJ, TIEFENBACH, P. Navigationssystem für mobilitätsbeeinträchtigte Menschen (NafmoM), in: STROBL, J. et al. (Eds.), *Angewandte Geoinformatik 2006. Beiträge zum 18. AGIT-Symposium Salzburg* (Heidelberg, 2006) pp. 597-602.

[2] STARK, HJ. OpenAddresses.ch, in: Tagungsband *Konferenz für Freie und Open Source Software für Geoinformationssysteme* (Freiburg i.Brs., 2008) pp. 55-60.

[3] STARK, HJ. OpenAddresses - reloaded. *Geomatik Schweiz* (05/2010).

[4] www.mapfish.org [accessed June, 15 2012]

[5] STARK, HJ. , KARRIE, C., BÄHLER, L. Schwachstellenanalyse im Langsamverkehr. in: STROBL, J. et al. (Eds.), *Angewandte Geoinformatik 2011. Beiträge zum 18. AGIT-Symposium Salzburg* (Heidelberg, 2011) pp. 421-427.

[6] STARK, HJ. , KARRIE, C. Anwendung im Bereich Bauzonenstatistik in der Schweiz, in: Tagungsband *Konferenz für Freie und Open Source Software für Geoinformationssysteme* (Dessau, 2012) pp. 46-50.

[7] STARK, HJ. Spatial Pattern Analysis of Secondary School Students' Leisure Behaviour. in: CAR, A. et al. (Eds.), *Geospatial Crossroads @ GI_Forum '12*. Proceedings of the Geoinformatics Forum Salzburg. (Heidelberg, 2012) [currently in print].

[8] STARK, HJ., TREUTHARDT, C. (2011): Switzerland: Introducing Geo-Sensor Technologies and Cartographic Concepts Through the Map your World Project. In: *International Perspectives on Teaching and Learning with GIS in Secondary Schools*. Springer. Dordrecht Heidelberg London New York. 2011.

[9] YOSHIDA, D., SONG, X. AND RAGHAVAN, V. Development of track log and point of interest management system using Free and Open Source Software. *Applied Geomatics* 2(3) (2010), 123-135

- [10] MODSCHING, M., KRAMER, R., GRETZEL, U. AND HAGEN, K. Capturing the Beaten Paths: A Novel Method for Analysing Tourists' Spatial Behaviour at an Urban Destination. *Information and Communication Technologies in Tourism*. M. Hitz, M. Sigala and J. Murphy (Springer Vienna 2006). pp. 75-86
- [11] SCHLICH, R., SCHÖNFELDER, S., HANSON, S. AND AXHAUSEN, K. W. Structures of Leisure Travel: Temporal and Spatial Variability.» *Transport Reviews* 24(2) (2004), 219-237
- [12] SHOVAL, N. AND ISAACSON, M. Application of Tracking Technologies to the Study of Pedestrian Spatial Behavior*. *The Professional Geographer* 58(2) (2006), 172-183.
- [13] SHOVAL, N. AND ISAACSON, M. Tracking tourists in the digital age. *Annals of Tourism Research* 34(1) (2007), 141-159
- [14] GOLUB, S. Heatmap. Retrieved 10.01.2012, from <http://sethoscope.net/heatmap/>

Hydroweb: an Open Source educational WebGIS platform for the understanding of spatio-temporal variations of meteorological parameters at the watershed scale

Authors

- **Stéphane Joost**, Ecole Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), GIS Research Laboratory (LASIG), *Switzerland*
- **Jens Ingensand**, INSER SA, *Switzerland*
- **Michael Kalbermatten**, Système d'Information du Territoire of the canton of Neuchâtel (SITN), *Switzerland*
- **Rolf Peter Tanner**, Pädagogische Hochschule PHBern, Institut Sekundarstufe I, *Switzerland*
- **Vincent Luyet**, Laboratory of Environmental Fluid Mechanics and Hydrology (EFLUM), School of Architecture, Civil and environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), *Switzerland*

KEYWORDS : WebGIS platform, primary and secondary education, watershed analysis, pedagogical notebook, SVG, PHP, PostgreSQL/PostGIS

Introduction

GIS is an interdisciplinary technology able to support high-level thinking and spatial reasoning. It allows students to visualize complex real world problems, and supports multiple modes of learning [1]. But GIS remain unused at the primary school level. Indeed, there is a gap between a declared interest in

GIS technology and their effective slow rate of implementation in classrooms [2]. The complexities of desktop GIS are possible brakes on their adoption in schools. They can be cut by using web-based GIS solutions. Indeed, internet-based mapping provides an invaluable way for establishing GIS technology in the primary and secondary education (“K-12 education community” in USA and Canada), while avoiding the main barriers associated with desktop GIS. Such tools can support standard methods of teaching and learning while providing basic analysis tools for studying and exploring geographic or other scientific data in the classroom [3]. This kind of platform is ideal for many teachers and 12-14 years old children that are not able to spend the time and energy required to run desktop GIS ([4], [5]).

Bodzin and Anastasio [6] have identified several internet-mapping solutions dedicated to earth and environmental systems education. Focusing on watershed science, a very interesting Web-based GIS tool was developed as part of a teacher training project in the State of Michigan that focused on Comparing and Contrasting the physical aspects and human use of Watersheds In Michigan (CCWIM, [1]). On this basis, a Web-based GIS application named the H2OMapper was developed to provide an online map-based environment for storing data on water quality, terrain, and land use for different watersheds in Michigan (<http://h2omapper.resa.net/pmapper/map.phtml>). The use of the platform was included in a specific program developed by the Michigan Department of Environmental Quality (DEQ) to help schools integrate specific economic and environmental materials into their science and social studies curriculum consisting of five foundation environmental education concepts: ecosystems and biodiversity, land use, water quality, energy resources, and air quality.

Given the substantial and growing need for materials development in the areas of environmental science and spatial thinking [7], we developed Hydroweb, an educational WebGIS platform for the understanding of spatio-temporal variations of meteorological parameters at the watershed scale dedicated to primary and secondary education in Switzerland (<http://lasigpc28.epfl.ch/hydroweb/>). This platform is part of a Swiss environmental education program named “CCES@School” conducted by the Competence Center Environment Sustainability (CCES) of the ETH domain, the union of Swiss Federal

universities and research institutions (<http://www.cces.ethz.ch>). CCES establishes a new thinking and priority setting based on sound scientific and engineering knowledge to integrate the principles of sustainable development into country policies and programs. One of its missions in particular is to achieve excellence in education and research and to focus research on crucial themes, ranging from climate and environmental changes to food safety, sustainable land use, natural resources and the management of natural risks.

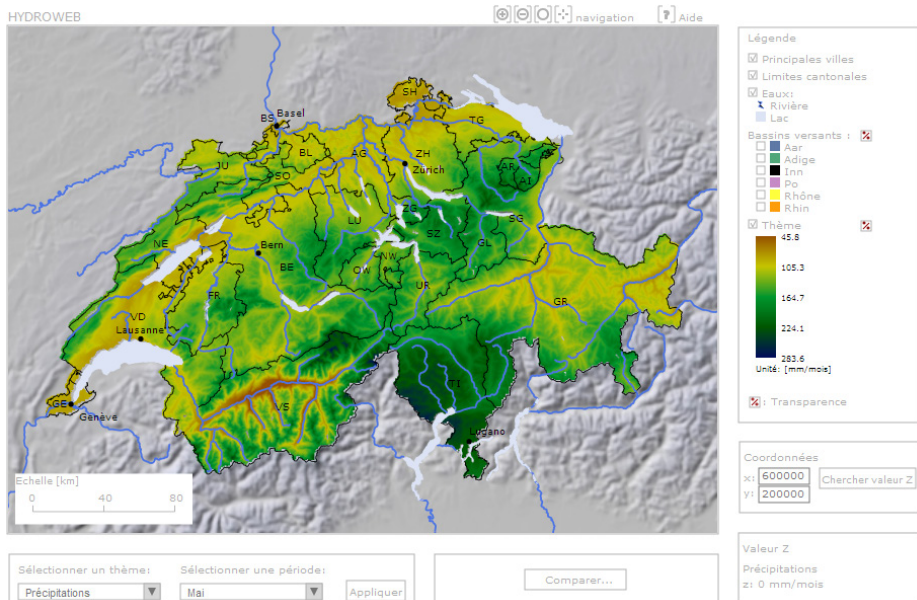
The project's main objective is to make Swiss students (12 to 14 years old children) aware of the spatial and temporal variations of environmental parameters at the watershed scale. It is also to make them understand the influence of the latter on the water balance as the result of the combination of several complex dynamic processes in which meteorological, soil, topography and anthropogenic factors are involved. Hydroweb also focuses on river discharge, a concrete impact of the water balance visible in the landscape. Finally, Hydroweb aims to stimulate the curiosity of children towards the functioning of their natural environment. Based on pedagogical concepts (inquiry-based and student-led investigations, open-ended questions, and real world experiences learning environment), Hydroweb is made of two main complementary modules.

WebGIS platform and data

The Hydroweb project was conducted by a consortium constituted of two laboratories of the Ecole Polytechnique Federale de Lausanne (EPFL) and of the Educational Science University of Bern (PHBern). Laboratories of Environmental Fluid Mechanics and Hydrology (EFLUM) and of Geographic Information Systems (LASIG) have developed the technical content of the WebGIS platform while the Educational Science University of Bern has elaborated the educational part of the project [8].

The WebGIS platform uses environmental information layers (see **Figure 1**). This tool is easy-to-use and permits to visualize, compare, look for and analyze monthly, seasonal and yearly meteorological data (temperature, precipitation and evaporation). Online instructions explain how to use the WebGIS platform.

FIGURE 1



The data constituting GIS layers (topography, state borders, watershed limits and hydrographic network) have been produced by swisstopo (www.swisstopo.ch). Meteorological data were recorded by a network of ground-based automatic weather stations operated by MeteoSwiss (<http://www.meteosuisse.ch>). A mean of Temperature, precipitation and evaporation values between 1961 and 1990 was calculated, generalized and made available in raster layers whose spatial resolution is 25 meters.

The challenge in the development of Hydroweb was to find an architecture that could easily be migrated to other servers in order to be used within an educational institutions server

infrastructure. This implied that we had to ignore map-engines such as MapServer, webservices such as WMTS and complex database engines and structures. We therefore chose to base Hydroweb on the open standard SVG both for the interface and map-rendering and javascript for making the system dynamic. These standards are natively supported in major browsers. The available layers are either stored in raster-images or, if vector-based, in SVG-shapes. Through the use of javascript, layers are dynamically loaded into the browser's memory. One further challenge was the display of map-related values in the interface (e.g. when the user moves his pointing device over the map). This functionality was implemented through the generation of value-matrices that are dynamically loaded at the same time a layer is loaded. Javascript dynamically interpolates values corresponding to both, the pointing device's screen coordinates and the layer and displays the calculated values in the system's interface.

Some parts of the SVG interface are dynamically generated by PHP (e.g. when the interface's language is changed). Therefore whole the system only needs a simple webserver and PHP as a minimum installation to run. For the pre-calculation of the value-matrices and histograms the well-known PostgreSQL/PostGIS database system was used.

Pedagogic notebook

A pedagogical notebook proposing applied educational activities grouped according to a) the variation of environmental parameters (temperature, precipitation and evaporation) and b) hydrological processes (river discharge, rain regime, watershed analysis). The exercises are based on the use of the WebGIS system and the students have to answer different questions (e.g. identify the geographical coordinates a specific location using a topographic map, extract precipitation and temperature data with Hydroweb, and create a seasonal climatogram for this place with the help of a spreadsheet). The different manipulations permit to develop specific competences like graph elaboration, measure reporting, data analysis, discussion and thinking about a given topic, etc. The notebook constitutes an initiation to scientific investigations; it proposes a questioning phase, the formulation of work hypotheses, the analysis of data and the discussion of results obtained. All

activities are independent. Some of them can spread over several months and other are punctual. All activities can be enriched, modified, or completed according to the scholar level, and adapted to specific objectives to be reached and to teaching conditions.

Conclusion

The Hydroweb platform allows students to manipulate georeferenced meteorological data through a very simple WebGIS interface, and to understand their geographical distribution during the different months or seasons in the various watersheds of the Swiss territory. Its architecture is based on open source GIS components and can easily be migrated to other servers in order to be used within any educational institution server infrastructure. Coupled with a teaching guide that includes thematic activities, Hydroweb is an open-ended and inquiry-based learning environment allowing pupils to lead their own investigations. It has direct correlation to real world experience, enables the visualization of complex real world problems, and supports multiple modes of learning.

[1] HENRY, P. and SEMPLE, H. Integrating Online GIS into the K–12 Curricula: Lessons from the Development of a Collaborative GIS in Michigan. *Journal of Geography*, 111, 1 (2011), 3-14.

[2] KERSKI, J. A nationwide analysis of the implementation of GIS in high school. In *Proceedings of the Nineteenth Annual ESRI International User Conference* (San Diego, California, 1999).

[3] BAKER, T. R. Internet-Based GIS Mapping in Support of K-12 Education. *The Professional Geographer*, 57, 1 (2005), 44-50.

[4] CHEUNG, R. L. and BROWN, S. Designing a distributed geographic information system for environmental education. In *Proceedings of the Twenty-First Annual ESRI International User Conference* (San Diego, California, 2001).

[5] O'DEA, E. *Integrating Geographic Information Systems and community mapping into secondary science education: A web GIS approach*. Oregon State University, 2002.

- [6] BODZIN, A. and ANASTASIO, D. Using Web-based GIS for earth and environmental systems education. *Journal of Geoscience Education*, 54, 3 (2006), 297–300.
- [7] BAKER, T. R., PALMER, A. M. and KERSKI, J. J. A National Survey to Examine Teacher Professional Development and Implementation of Desktop GIS. *Journal of Geography*, 108, 4-5 (2009), 174-185.
- [8] LUYET, V., JOOST, S., JUD, S., TANNER, R. P. and AUER, M. *Hydroweb*. The CCES Environmental Education Initiative, ETH, Zürich, 2011.

Establishing a US National FOSS4G Academy

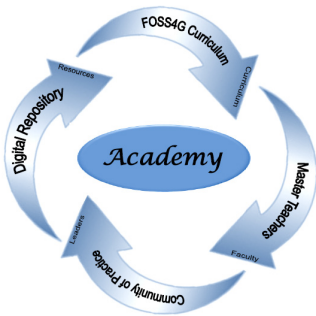
Author

Phillip Davis, US National Geospatial Technology Center of Excellence (GeoTech Center), *United States*

KEYWORDS : higher education, FOSS4G, GTCM, GeoTech, MOOC, flipped learning, academy, center of excellence, reform, DOL, curriculum

Motivating Rationale

The **Free & Open Source Software for Geospatial (FOSS4G) Academy** project will create curriculum and educational materials and perform professional development under a grant from the US National Science Foundation. The Academy will consist of four components: a) certificate program based on seven FOSS4G courses and associated material, b) cadre of 50 geospatial faculty participants nationwide, educated in the latest FOSS4G software and Universal Learning by Design [4, 21] pedagogy practices, c) state-of-the-art open source eLearning platform capable of engaging 1000s of simultaneous learners and incorporating the latest in learner performance analytics and feedback, and d) community of practice, including extensive use of open source collaboration tools and social media to connect and sustain the Academy members over the long-term. The resulting FOSS4G curriculum will lead to a well-defined career pathway [24] in geospatial technology that can be offered at the undergraduate level as either a certificate or basis of an AAS or BS degree. It will incorporate the latest in pedagogy research and practice including the use of “flipped” lessons, massive open online courses (MOOCs) [1, 8, 25] and micro credentials [2, 3] to increase the quality and quantity of geospatial technology workers in the coming decade. *The*



*Vision of the Academy is to establish a vibrant community of educators, learners, and users of Free & Open Source Software for Geospatial applications. The **Mission** of the Academy is to create and sustain this vibrant community of educators dedicated to promoting the use of Free & Open Source Software for Geospatial (FOSS4G) applications to help meet America's need for a well educated Geospatial Technology workforce.*

Need for Academy

The Geospatial Technology Industry is a high-growth, high-technology field [6, 12] first recognized by the DOL in 2005 as being essential to America's economic recovery and national security. The industry was formally defined in June, 2010 with the publication of the Geospatial Technology Competency Model (GTCM) on the DOL Competency Model Clearinghouse. The industry is projected to require an additional 150,000 US technicians between 2010 and 2020 [6], at which time total employment will increase from the current 447,000 US workers to approximately 600,000 [6].

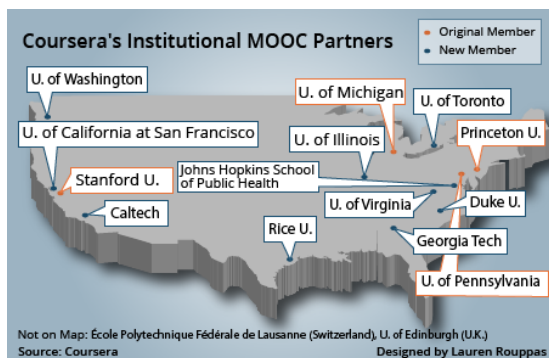
Occupation	Employment (2010)	Projected Growth (2010-2020)	Projected Growth Rate (2010-2020)
Geospatial Information Technician	210,000	51,600	3 to 9%
Remote Sensing Scientists and Technologists	30,000	13,300	3 to 9%
Remote Sensing Technicians	62,000	33,500	10 to 19%
Geodetic Surveyors*\	51,000	24,200	20 to 28%
Mapping Technicians	57,000	20,000	10 to 19%
Cartographers and Photogrammetrists	14,000	6,100	20 to 28%
Totals	424,000	148,700	(3 to 28%)

TABLE 1 - Project Geospatial Job Grow (DOLETA)

Source : U.S. Department of Labor Employment and Training Administration, O*NET Online, <http://online.onetcenter.org/>, August 1, 2012

Geospatial Occupations U.S. Department of Labor Employment and Training Administration

These workers will need the broadest range of knowledge, skills, and abilities (KSA) to sustain them throughout their professional lifetimes. The GTCM offers a comprehensive list of these KSAs, and describes the need for workers capable of using a variety of tools and technologies. The US higher education system currently offers geospatial technology in approximately 800 colleges and universities [18]. 750 of these colleges report offering geospatial software exclusively from a single proprietary, closed-sourced, vendor. The lack of variety in geospatial application software on these campuses limits the student's technology experience. What students require is a broad and inclusive range of technology exposure. This project will build a robust, GTCM-aligned, curriculum that will utilize a wide-range of FOSS4G software applications [14, 15] to broaden learner's KSA in a multiplicity of alternative tools.



Intellectual Merit

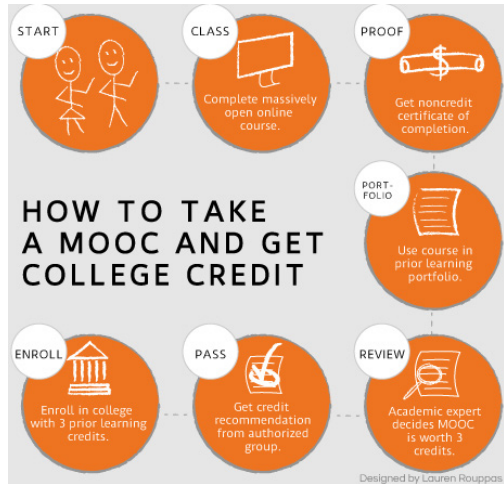
The Academy will 1) incorporate research-based instructional design strategies [7, 8, 11, 21] that build upon cognitive science and existing resources developed by leading-edge institutions, 2) it will incorporate potentially transformative strategies,

like the new flipped classroom [13, 20, 23] demonstrated successfully by Khan Academy and pioneered by the new generation of online course providers like Udacity and Coursera [1, 3, 8, 23]. The Academy will 1) introduce the students to a new, community based software development, user support and business model [2, 3, 7] that lowers the barriers for entrepreneur initiative. The FOSS4G Academy will 2) bring innovation into higher education by providing a deeper insight into the technology development and tools implementation enabled by the open source code and open participation in the development process [5, 9, 10, 17].

Broader Impact

The Academy incorporates 1) an innovative micro-credentialing plan [2, 3, 7] to allow learners to receive formal recognition of their new knowledge and skills. Through the use of authenticated badges, learners can accumulate digital evidence of their knowledge, skills, and abilities and may receive transferrable academic credit through a network of partner colleges. We will 2) develop a Spanish language version of the FOSS4G curriculum appropriate

for the young Spanish-speaking adult immigrants, primarily from Mexico, identified by the Obama administration's recently-passed US DREAM Act [16]. This act has the potential to add an additional 800,000 young working-aged adults (under 30), along with another 350,000 middle- and high-school dropouts to the US workforce.



[1] A MASSIVE OPEN ONLINE COURSE. n.d. Retrieved July 17, 2012 from <http://gtmooc.com/>

[2] AN OPEN BADGE SYSTEM FRAMEWORK: A foundational piece on assessment and badges for open, informal and social learning environments. 2012. DML Central. Retrieved May 22, 2012 from <http://dmlcentral.net/resources/4440>

[3] BATES, T. 2012. The challenge of converting MOOCs (or anything else) into college credits. Online Learning and Distance Education Resources. Retrieved August 10, 2012 from <http://www.tonybates.ca/2012/08/08/the-challenge-of-converting-moocs-or-anything-else-into-college-credits/>

[4] CREATIVE COMMONS. n.d. Retrieved May 9, 2012 from <http://creativecommons.org/>

- [5] DAVIS, J. Mil-OSS. Retrieved May 1, 2012 from <http://www.mil-oss.org/>
- [6] DEMPSEY, C. 2012. Where are the GIS Jobs? A Look at the GIS Job Market in the United States. GIS Lounge. Retrieved June 28, 2012 from <http://gislounge.com/where-are-the-gis-jobs-a-look-at-the-gis-job-market-in-the-united-states/>
- [7] FAIN, P. 2012. Making it Count. Inside Higher Ed. Retrieved August 6, 2012 from <http://www.insidehighered.com/news/2012/06/15/earning-college-credit-moocs-through-prior-learning-assessment>
- [8] FERENSTEIN, G. Move Over Harvard And MIT, Stanford Has The Real Revolution In Education. Mid-Pacific ICT Center. Retrieved August 2, 2012 from <http://mpictcenter.blogspot.com/2012/06/move-over-harvard-and-mit-stanford-has.html>
- [9] MENKE, K. 2011. FOSS4G Workshop for Educators at FOSS4G. Bird's Eye View GIS. Retrieved May 1, 2012 from <http://www.birdseyeviewgis.com/blog/2011/7/13/foss4g-workshop-for-educators-at-foss4g.html>
- [10] GAKSTATTER, E. 2012. Open Source GIS: Part II. Geospatial Solutions. Retrieved July 19, 2012 from <http://www.gpsworld.com/gis/gss-weekly/open-source-gis-part-ii-13011>
- [11] HILL, P. 2012. The Emerging Landscape of Educational Delivery Models. eLiterate. Retrieved April 10, 2012 from http://mfeldstein.com/the-emerging-landscape-of-educational-delivery-models/?utm_source=facebook&utm_medium=social&utm_content=bd588b63-b050-407d-b8ff-8a61c55b217d
- [12] INDUSTRY OUTLOOK 2012: The Year of 'The Cloud?' (Or Did We Just Have It?). 2012. Geoplace.com. Retrieved August 8, 2012 from <http://www.geoplace.com/ME2/dirmod>.
- [13] KAMENETZ, A. 2012. How Coursera, A Free Online Education Service, Will School Us All. Fast Company. Retrieved August 10, 2012 from <http://www.fastcompany.com/3000042/how-coursera-free-online-education-service-will-school-us-all>
- [14] KOUYOUJIAN, V. 2011. Open Source, Closed Source: Moving to the Middle. Esri. Retrieved April 13, 2012 from <http://blogs.esri.com/esri/esri-insider/2011/10/24/open-source-closed-source-moving-to-the-middle/>
- [15] LARGE TURN-OUT DEMONSTRATES GOVERNMENT'S INTEREST IN OPEN SOURCE SOFTWARE. 2012. Geospatial Solutions. Retrieved June 29, 2012 from <http://www.gpsworld.com/gis/geointelligence/news/large-turn-out-demonstrates-governments-interest-open-source-software-13107>

- [16] LOVE, J. 2012. U.S. Opens a Door to a Dream for Young Illegal Immigrants. The Chronicle of Higher Education. Retrieved August 15, 2012 from <http://chronicle.com/article/US-Opens-a-Door-to-a-Dream/133649/>
- [17] MCCORMICK, C. 2011. Conversations about open source and higher education at FOSS4G. Spatial Networks. Retrieved July 11, 2012 from <http://spatialnetworks.com/blog/2012/04/conversations-about-open-source-and-higher-education-at-foss4g/>
- [18] MENKE, K. 2011. Experiences Teaching Free and Open Source GIS at the Community College Level. Directions Magazine. Retrieved May 3, 2012 from <http://www.directionsmag.com/articles/experiences-teaching-free-and-open-source-gis-at-the-community-college/212522>
- [19] MEYERS, R. 2012. Coastal Bend colleges, universities recognized for graduating Hispanics in high-demand fields. Caller.com. Retrieved August 10, 2012 from <http://www.caller.com/news/2012/aug/02/area-colleges-and-universities-recognized-for-in/>
- [20] PARRY, M. 2012. Supersizing the College Classroom: How One Instructor Teaches 2,670 Students. The Chronicle of Higher Education. Retrieved August 8, 2012 from http://chronicle.com/article/How-One-InstructorTeaches/131656/?sid=pm&utm_source=pm&utm_medium=en
- [21] STANSBURY, M. 2012. Universal Design for Learning: The next big thing in school reform? eSchool News. Retrieved May 18, 2012 from <http://www.eschoolnews.com/2012/05/16/universal-design-for-learning-the-next-big-thing-in-school-reform/>
- [22] STOJIC, M. 2012. Are Defense Cuts Good for the Geospatial Sector? Directions Magazine. Retrieved May 3, 2012 from <http://www.directionsmag.com/articles/are-defense-cuts-good-for-the-geospatial-sector/247070>
- [23] STOMMEL, J. 2012. The March of the MOOCs: Monstrous Open Online Courses. HYBRID PEDAGOGY: A Digital Journal of Teaching & Technology. Retrieved July 31, 2012 from http://www.hybridpedagogy.com/Journal/files/MOOC_MOOC.html
- [24] UN COMMITTEE OF EXPERTS ON GLOBAL GEOSPATIAL INFORMATION MANAGEMENT. Future trends in geospatial information management: the five to ten year vision. n.d. UN-GGIM. Retrieved June 10, 2012 from http://ggim.un.org/docs/meetings/Netherlands/Future%20trends%20in%20geospatial%20information%20management%20summary_12April.pdf
- [25] WHAT IS A MOOC? 2010. YouTube. Retrieved August 6, 2012 from <http://www.youtube.com/watch?v=eW3gMGqcZQc>

Design and implementation of a distance education course on open source web mapping

Authors

- **Barend Köbben**, ITC - University of Twente, Faculty of Geo-Information Science and Earth Observation, *Netherlands*
- **Ivana Ivanova**, ITC - University of Twente, Faculty of Geo-Information Science and Earth Observation, *Netherlands*

KEYWORDS : open source, distance education, web mapping, didactics

1. Introduction

In this paper, we discuss the design principles used in the development and implementation of the *Distance Education course on Open Standards and Open Source Web mapping* (DE-OSM). The main objective of this course is to teach participants about the principles of using open standards for web mapping using open source tools. More specifically, participants will learn :

- the principles of interoperability and geo-webservices
- the principles of open standards in the geo-domain
- the principles of client-server architecture in the geo-domain
- to deploy open standards compliant geo-webservices using open source tools
- to build browser-based clients using open source software

The DE-OSM course was developed at ITC, the Faculty of Geo-Information Science and Earth Observation of the University of Twente. ITC is an institute that aims at capacity building and institutional development in various fields that include the use of Earth observation, collection and management of spatial information, and the development of data integration methods.

2. Design of the course

The DE–OSM course is designed as a *self-instructive distance education* course. The main reason for choosing a *distance education* format is to offer students a learning environment that can be accessed in *their own time*, and can be processed at *their own pace* in *their own location*.

The course being *self-instructive* means that the content of the course itself acts as a teacher. This may sound obvious, but it is quite difficult to apply sound instructional principles in content material for ‘distant’ students, as compared to ‘face-to face’ students. The material in any lesson must explain itself. It also has to provide everything necessary for the student to behave in a self-instructive manner, and to gather the knowledge required. To reach these goals, the DE–OSM course applies the didactic approach known as the ICARE (Introduce, Connect, Apply, Reflect, Extend) system, supported by the principles of instruction theory [9]. ICARE is a five-part system that borrows aspects from various taxonomies, including Gagné’s Nine events of instructions [3], Merrill’s Component display theory [8], and Bloom’s Taxonomy of higher order learning [1].

The DE–OSM course is offered as a set of stand-alone course components, accessible in an order as preferred by the student. The course set-up and structure naturally suggest an ideal path through the material, in which a lecture is the central course component. A lecture acts as the ‘knowledge glue’ that helps the students in constructing their new knowledge. This approach arises from the characteristics of the topic (fundamentals of open standards and open source web mapping), since it introduces principles (axioms, theorems, and base skills). The expected entrance knowledge level is low; therefore, it is less suitable to allow students (who do not yet master the principles) to formulate their own opinions through the application of collaborative pedagogical methods. Collaborative learning techniques, such as discussions and collaborative presentation, are more suitable for advanced students and we apply them towards the end of courses.

The content of the course is self-explanatory, meaning that the student should find everything he or she needs for comprehension of the course content in the provided materials. This approach allows the teacher to act more as a tutor than a sage during the course. This significantly decreases the staff time required

to run the course. Although the time invested by the tutor during the course is diminished, the students' performance remains of interest to the tutor. Discussions will be stimulated and must be carefully monitored. Depending on the length of a course, the content may be divided into units and every unit consists of lessons, where the lesson is the main Learning Object (LO). The term "Learning Objects" started to find its way into the world of education at the end of the 1990s. LOs are "instructional components that can be reused a number of times in different learning contexts" [10]. Learning objects are the most basic building block of a lesson or activity, and they are :

- searchable
- usable in any learning environment
- capable of being grouped or to stand alone
- transportable from course to course and program to program

In our course design, we adopted Chiappe's definition, in which an LO is "a self-contained, reusable entity with a clear learning aim that contains at least three internal changing and editable components: content, instructional activities (learning activities), and context elements" [2]. We chose the lesson as the main learning object. It has clearly stated objectives, has well-defined and limited content, and can be assessed. All elements used to build such an LO are seen as course components.

Lessons in ITC's distance courses typically have six types of components: [1] introduction (in the form of a study guide), [2] lecture(s), [3] demonstration(s), [4] exercise(s), [5] self-test(s) and [6] a reader or book. A general overview of the didactical structure of distance education courses at ITC is provided in Figure 1. Some of these components are not independent, and only make sense in combination with their specific reference component (e.g. a demonstration that explains a certain action described in the lesson). Other components can easily be reused in other lessons in another context (e.g. a reader or exercises).

Of course, any course design should also take into account the skills and background of the potential participants. In our case, the target audience is primarily the staff of European national mapping and cartographic agencies and other organizations with a need for using open source web mapping. Advanced students in the geo-domain that want to apply open standards and open source web mapping in their studies are also considered as part of the target group.

We expect these participants to be advanced learners, with a professional understanding of the underlying principles of geo-data.

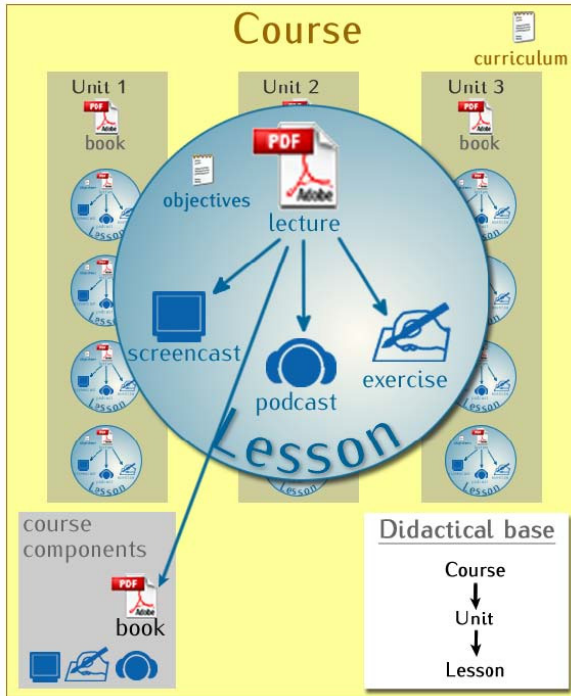


FIGURE 1

Didactical structure used in distance education courses at ITC. Note that not all elements are necessarily used in all lessons (e.g. the DE–OSM course has no pod- and screencasts)

In the paper, we will describe in detail how we applied the didactic principles and the distance education framework we developed at ITC for the particular target group of the two week DE–OSM course.

3. Implementation of the course

The course is accessible through a Learning Management System (the current version of the course runs on the Moodle platform), requiring only a web browser on the student's end. This LMS allows students to interact with tutors, by means of a discussion board. Since the LMS client contains an automatic update mechanism, students will always access the most up-to-date version of the course. The elements of the courseware (reader, lectures,

exercises) are offered as interactive PDF files. The DE–OSM course provides the participants with knowledge and tools to set up a web mapping application according to the principles of what we call **SDI^{light}**. The term SDI for Spatial Data Infrastructure may be usually connected with (very) large regional or national spatial data warehouses, but the principles of SDI can also be applied in more simple and cost–effective ways. The down–to–earth approach of SDI^{light} provides students with a platform for relatively simple, low–cost, yet powerful ways of sharing data amongst various distributed offices and institutions as well as the general public. To achieve this, we use open standards whenever available and open source solutions where possible. More on the SDI^{light} approach can be found in [7].

SDI^{light} building blocks are open source tools and include :

- a PostgreSQL/PostGIS spatial database back–end that stores the spatial data using the Open Geospatial Consortium Simple Features specifications
- MapServer, the server applications that allows dissemination of spatial data over the Internet, which is fundamental requirement for web mapping
- simple browser-based clients enabling access to the required maps and data. At present, we employ various techniques such as dynamic HTML, eXtensible Markup Language, asynchronous JavaScript (using the OpenLayers application programming interface) and the Scalable Vector Graphics image format

3.1 Lecture - the knowledge glue

In our course, the lecture acts as the *knowledge glue*. The lecture draws together ideas presented in other elements of the courseware (reader, demonstration, exercise, self-test). We typeset most of our courseware using LATEX, a high-quality typesetting system based on TEX, developed by Donald Knuth in 1978 [6]. The scripting language used in the course (JavaScript and HTML) has a special flavour, by which we mean native code typesetting and use of colors in most interpreters is catered for by using a special styling in LATEX, namely the modified `listing` environment. Authors of the lecture can simply copy code snippets from the interpreter and paste them into the `.tex` source file of the lecture.

All that is required is to insert a piece of code into the .tex file of the lecture such as the following :

```
begin{lstlisting}
  Map types produced by Lovely Maps:
  <ol>
    <li>tourist maps</li>
    <ol type=i >
      <li>monochrome </li>
      <li>colour</li>
    </ol>
    <li>location maps</li>
    <li>route maps</li>
    <li>imaginary maps</li>
  </ol>
\end{lstlisting}
```

The above example will be typeset as shown in **Figure 2**. As mentioned earlier, the DE–OSM lecture didactical system sees the lecture as *knowledge glue*. This is exemplified by a multimedia strip on the right margin of the lecture page. Every page of the lecture can point to a (video) demonstration, an exercise, and/or a specific location in the reader.

3.2 Software installation

Our technical realization reduces the initial technicalities related to installation and optimization of the work environment. This allows students and tutors to concentrate on the principles of web mapping and leave the installation details for later. The LMS client interface also allows students to interact with tutors, by following the discussion board. Since the LMS client contains an automatic update mechanism, students will always access the Design and implementation of a distance education course on open source web mapping 5 **Figure 2**. Lecture in DE–OSM most up-to-date version of the course.

Web mapping technology required by the content of the course is installed on ITC's servers and is accessible to the students. Once the course is finished, students will receive installation instructions to be able to replicate the set-up

Introduction
 Before we start coding
 HTML basics
 HTML content types

Text
Lists

Unordered list

Ordered list

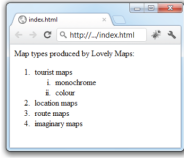
Example:


```

Map types produced by Lovely Maps:
<ol>
<li>tourist maps</li>
<ol type=i >
<li>monochrome </li>
<li>colour</li>
</ol>
<li>location maps</li>
<li>route maps</li>
<li>imaginary maps</li>
</ol>

```

... and the result:




Introduction to HTML

page 51 of 49

FIGURE 2

Lecture in DE-OSM

at their own location. Although the nature of studying of a scripting language and web application building is through self-study, this client-server solution also allows components of the course to be deployed in a face-to-face setting, as well as in the distance education mode.

3.3 DE course execution

The duration of DE-OSM course is two weeks. We expect that students will study parttime. As our previous experience shows, studying part-time means perhaps a few hours in the evening, but mostly during the weekend. This typical study behavior strengthens our choice for self-instructive nature of the course.

Assessment is twofold :

1. testing the theoretical knowledge with tests conducted in LMS
2. testing the practical knowledge through an assignment to build a web map application

Testing will be scheduled only at the end of the course. We allow the students submitting their web map applications one month after the course.

4 Results

One of the key objectives in the design was to lower the need for tutors' input during the course, without letting the student drown in lecture notes, readers, articles and instructions. We expect the design principles behind this course to ensure that this objective can be met, because they have been successfully applied in earlier, comparable, courses (e.g. [4], [5]).

At the time of writing this extended abstract, a first group of 24 participants has just finished the first run of the two week DE-OSM course. We can at this moment conclude that they by and large have been very actively involved and have finalized the course with the expected achievement : They have produced a working web mapping application according to the specifications given to them.

In the paper, we will look more closely to the course results as well as data we are currently gathering from the participants, to give a critical evaluation of the success of our approach.

Acknowledgments

The DE-OSM course was offered as part of the EduServ10 series (<http://euroedr.net/eduserv10>) and its development was supported by EuroSDR funding.

[1] BLOOM, B. S. *Taxonomy of educational objectives*. Longman, New York, NY, USA, 1956.

[2] CHIAPPE LAVERDE, A., SEGOVIA CIFUENTES, Y., AND RINCÓN RODRÍGUEZ, Y. Toward an instructional design model based on learning objects. *Educational Technology Research and Development* 55, 6 (2007), 671–681.

[3] GAGNÉ , R. M. *The conditions of learning.*, 4 ed. Rinehart andWinston Hold, New York, NY, USA, 1977.

- [4] IVÁNOVÁ, I., DE BY, R. A., BAKKER, W. H., KOOISTRA, J. F., FERINGA, W. F., AND RETSIOS, B. Principles of databases : distance education., 2010.
- [5] IVÁNOVÁ, I., HUISMAN, O., DE BY, R. A., FERINGA, W. F., BAKKER, W. H., RUTZINGER, M., AND RETSIOS, B. Programming skills: distance education., 2010.
- [6] KNUTH, DONALD, E. *The TeXbook*. Adison-Wesley, Reading, Massachusetts, USA, 1984.
- [7] KÖBBEN, B., FOERSTER, T., LEMMENS, R., BY, R. A. D., HUISMAN, O., AND MORALES, J. Using the SDlight approach in teaching a geoinformatics master. *Transactions in GIS* 14, s1 (2010), 25–37.
- [8] MERILL, D. M. Instructional transaction theory: Instructional design based on knowledge objects. *Educational Technology* 36, 3 (1996), 30–37.
- [9] PASTOR, N. The ICARE system. In *Encyclopedia of Educational Technology*. San Diego State University, 2000.
- [10] WILEY, D. A. Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In *The Instructional Use of Learning Objects*, D. A. Wiley, Ed. Association for Educational Communications and Technology, 2001, pp. 1–35.

r.rotstab: a GRASS-based deterministic model for deep-seated landslide susceptibility analysis over large areas

Authors

- **Martin Mergili**, Institute of Applied Geology, BOKU University of Natural Resources and Life Sciences Vienna, *Austria*
- **Ivan Marchesini**, CNR IRPI, Perugia, *Italy*
- **Mauro Rossi**, CNR IRPI, Perugia and Dipartimento di Scienze della Terra, Università degli Studi di Perugia, *Italy*
- **Fausto Guzzetti**, CNR IRPI, Perugia, *Italy*
- **Wolfgang Fellin**, Division of Geotechnical and Tunnel Engineering, University of Innsbruck, *Austria*

KEYWORDS : cloud computing, Collazzone area, factor of safety, GRASS GIS, slip ellipsoid, slope stability model

Landslides can be studied at site scale (i.e. for individual slopes) or at large (local) or small (regional) scales [1,2]. The deterministic approaches to the landslide stability assessment at site scales are quite classical and commonly accepted procedures. The deterministic analysis of landslides at local and regional scales (distributed models), and the relative estimation of landslide susceptibility, are non-standard tasks.

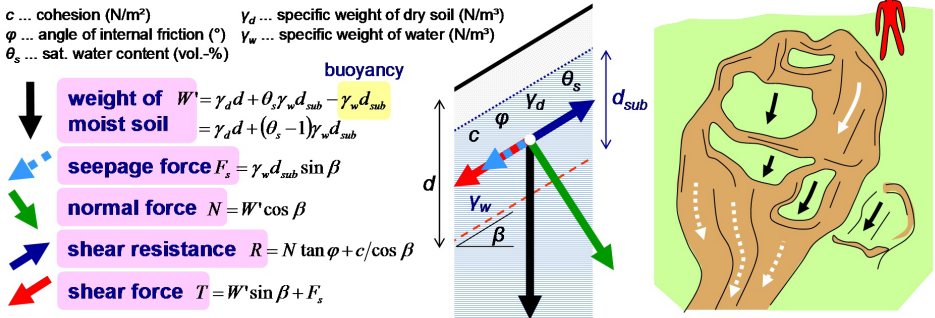
Deterministic slope stability models, based on limit equilibrium analysis, are applied to particular landslide types (e.g., shallow soil slips, debris flows, rock falls), or to investigate the effects of specific triggers, i.e., an intense rainfall event or an earthquake. In particular, distributed infinite slope stability models are used to evaluate the spatial probability of shallow slope failures.

In these models, the factor of safety is computed on a pixel basis, assuming a slope-parallel, infinite slip surface. They can be easily implemented in GIS environments since they do not rely on complex neighborhood relationships.

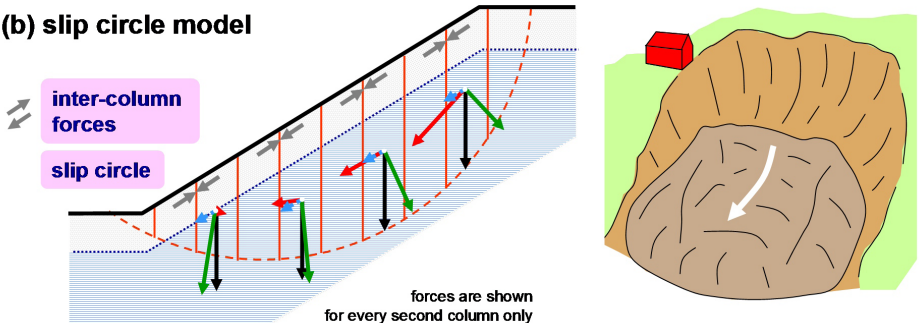
However, since shallow slope failures coexist locally with deep-seated landslides, infinite slope stability models fail to describe the complexity of the landslide phenomena. Limit equilibrium models with curved sliding surfaces are geometrically more complex, and their implementation with raster-based GIS is a challenging task. Only few attempts were made to develop GIS-based three-dimensional applications of such methods [3].

FIGURE 1

(a) infinite slope stability model unit raster cell size: 1 x 1 m slope-parallel seepage



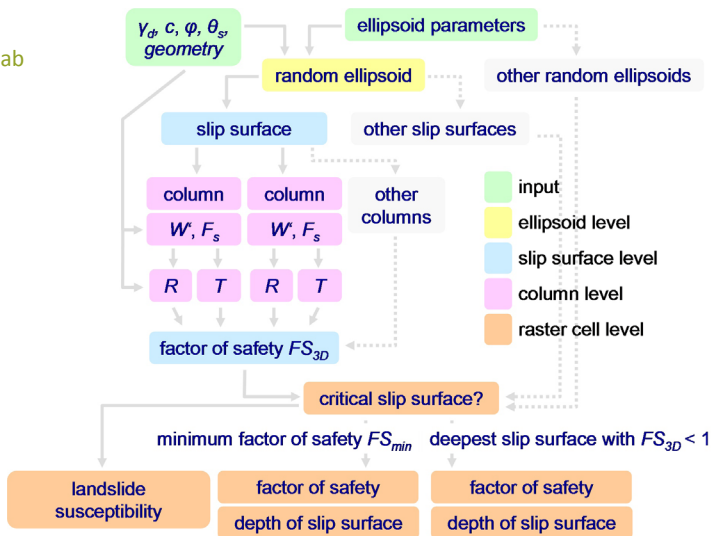
(b) slip circle model



We present a preliminary implementation of a GRASS GIS-based, three-dimensional slope stability model capable of dealing with both shallow and deep-seated slope failures. The Open Source GIS package GRASS GIS [4] environment offers comprehensive opportunities for spatial analysis, particularly raster operations. Simple analyses or standard procedures can be performed using the existing tools or combining them by shell scripting. More complex tasks or non-standard procedures can be performed by implementing new modules, making use of the Python or C languages.

FIGURE 2

Work flow of r.rotstab

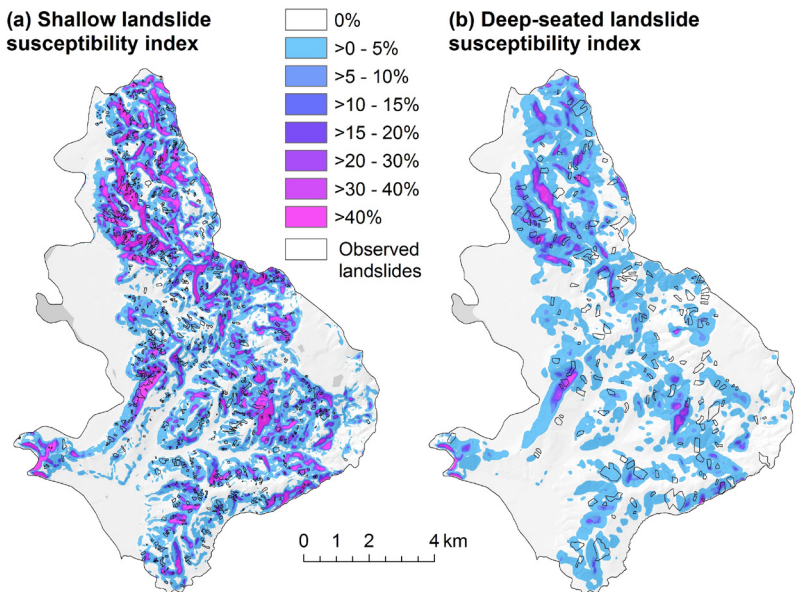


The model is developed and evaluated in GRASS GIS 6.4 version as the C-based raster module r.rotstab. Data management is facilitated by the shell script r.rotstab.sh. For large study areas the program includes the option to split the area into a number of tiles, to run the computation separately for each tile and at the end to combine the results for each tile (script r.rotstab.xl.sh). This avoids running into troubles with limited memory and allows to largely rely on ordinary arrays instead of segmentation files, which would considerably slow down the computation process. There has to be an overlap of the maximum extent of one ellipsoid between the tiles in order to avoid poorly covered areas at the edges.

The model makes use of a slight modification of the three-dimensional sliding surface model proposed by Hovland [5] and revised and extended by Xie et al. [3]. Given a Digital Elevation Model (DEM) and a set of thematic layers, mainly concerning geotechnical and hydraulic parameters, the model evaluates the slope instability over a large number of randomly determined potential ellipsoidal slip surfaces. In addition to ellipsoidal slip surfaces, truncated ellipsoids can be used to simulate the presence of shallow weak layers, delimited by soil discontinuities or hard bedrock. Any raster cell may be intersected by various sliding surfaces, each associated with a computed factor of safety. The lowest value of the factor of safety is stored for each raster cell together with the depth of the associated slip surface. This results in an overview of potentially unstable regions without showing the individual sliding areas. In addition, a landslide susceptibility index in the range 0 - 1 is provided, relating the number of unstable slip surfaces to the total number of slip surfaces simulated in each pixel.

FIGURE 3

Landslide susceptibility indices for shallow and deep-seated landslides



We test the model in the Collazzone area, Umbria, Central Italy, which is susceptible to landslides of different types. The presence of both shallow translational and deep-seated rotational landslides and the availability of reference data allow for the critical evaluation of the model in comparison with infinite slope stability models. For the calculation the entire Collazzone area is split into 150 tiles of approx. 1.7 x 1.6 km and 500,000 ellipsoids were simulated for each tile.

Exploiting slip surfaces truncated at a depth of 1.3 m - the average depth of shallow landslides recorded in the area - the model successfully predicts the observed landslides patterns. As expected for this type of landslides, the results are in general very similar to those yielded with the infinite slope stability model. However, the results yielded with *r.rotstab* are more smoothed since small-scale variations of topography (particularly slope) are smoothed out. Both models result in a significant number of false positive raster cells (i.e. stable cells, wrongly predicted as unstable), which may either be areas potentially affected by landslides in the future, or mispredictions due to insufficient parameter knowledge. Tuning the geotechnical parameters towards a lower number of false positives can only be done at the cost of an increased number of false negative raster cells, which are certainly mispredictions.

Whilst this problem is moderate for shallow landslides, it is more pronounced for deep-seated landslides reaching a maximum depth of 20 m in the Collazzone area. In that case, the distribution of the observed landslides is very likely to be conditioned also by factors not used - or not accounted for appropriately - in the model.

Since *r.rotstab* is designed specifically for this type of landslides, one of the main future tasks will be to explore further key parameters influencing deep-seated slope stability. For example the bedding attitude of the geological layers is supposed to be one of the most important aspects that condition the slope stability. According to field observations in the Collazzone area, morpho-

structural settings play a crucial role for deep-seated landslide distribution. A second key for improving the prediction rate will be to refine the knowledge on the spatial (particularly vertical) structure of the regolith parameters.

Besides the collection and preparation of additional data, more advanced and extended parameter tests will also be required to improve the model performance. Using a Virtual Machine (1 core - 3.00GHz, with 8 Gb of RAM) running Ubuntu 11.04, the current implementation is able to process several tens of millions ellipsoids per day. Further enhancement of the model performance will require the parallelization of the code in order to run in multiple core environments or on a grid infrastructure (cloud computing). For this purpose we intend to test the module r.cloud which will be part of the prospective release of GRASS 7 in the near future.

[1] VAN WESTEN, C.J. The modelling of landslide hazards using GIS. *Surveys in Geophysics* 21 (2000), 241-255.

[2] GUZZETTI, F. *Landslide Hazard and Risk Assessment*. Ph.D. Thesis, University of Bonn, Germany, 2006.

[3] XIE, M., ESAKI, T., QIU, C. AND WANG, C. Geographical information system-based computational implementation and application of spatial three-dimensional slope stability analysis. *Computers and Geotechnics* 33 (2006), 260-274.

[4] GRASS DEVELOPMENT TEAM 2012. *Geographic Resources Analysis Support System (GRASS) Software*. Open Source Geospatial Foundation Project. Retrieved September 5, 2012 from <http://grass.osgeo.org>.

[5] HOVLAND, H.J. Three-dimensional slope stability analysis method. Journal of the Geotechnical Engineering Division. *Proceedings of the American Society of Civil Engineers* 103 (1977), 971-986.

Cluster analysis of geological point processes with R free software

Authors

- **Marj Tonini**, University of Lausanne, *Switzerland*
- **Antonio Abellán**, University of Lausanne, *Switzerland*
- **Andrea Pedrazzini**, University of Lausanne, *Switzerland*

KEYWORDS : landslides, spatial point pattern, R free software, Ripley's K-function, Nearest Neighbour clutter removal

Introduction

Landslides, as many other geological events (e.g. earthquakes, volcanoes, etc), are normally not randomly distributed but grouped in clusters both in space and in time. The analysis of their spatial properties and distribution is fundamental to understand their predisposing factors, and for prevention and forecasting purposes. From a statistical point of view, geologic events can be represented as «point processes» whose spatial distribution can be analysed using mathematical models for irregular or random point pattern.

Pattern recognition and specifically cluster analysis includes algorithms aiming at grouping objects showing similar properties into the respective categories. Spatial clusters can be identified whenever the observed distance among groups of point locations in space is lower than the expected distance for a random distribution. This assumption can be accepted or rejected based on the results of statistic tests. For geological events, which intensity is not constant and clusters are often not isolated, their detection is not evident. A vast literature exists on the spatial analysis of landslide distribution, especially for susceptibility map purpose, [4,7,11,12,13,16] while their spatial characterisation by means cluster algorithms is less investigated [18].

Two main types of spatial cluster algorithms can be outlined: 1) global cluster indicator, allowing to measure and test for the randomness of the point process; 2) local cluster algorithms allowing to identify clusters in space and/or in time. The present study illustrates two examples of application of the two abovementioned kinds of analysis by means of two distinct case studies: 1) the Ripley's K-function was applied to assess the global spatial attraction (clustering) among mapped landslides; 2) Nearest neighbour clutter removal method was applied to automatically detect landslides in LiDAR (Light Detection and Ranging) point clouds.

Computations were carried out using R free software for statistical computing and graphics [14]. R is a free software environment integrating facilities for data manipulation, calculation and graphical display. The R base can be extended via packages available through the Comprehensive R Archive Network (CRAN) which covers a very wide range of modern statistics. More specifically, the spatial point pattern analyses of the geological events considered in the present study and their cluster detection were supported by the package spatstat [3].

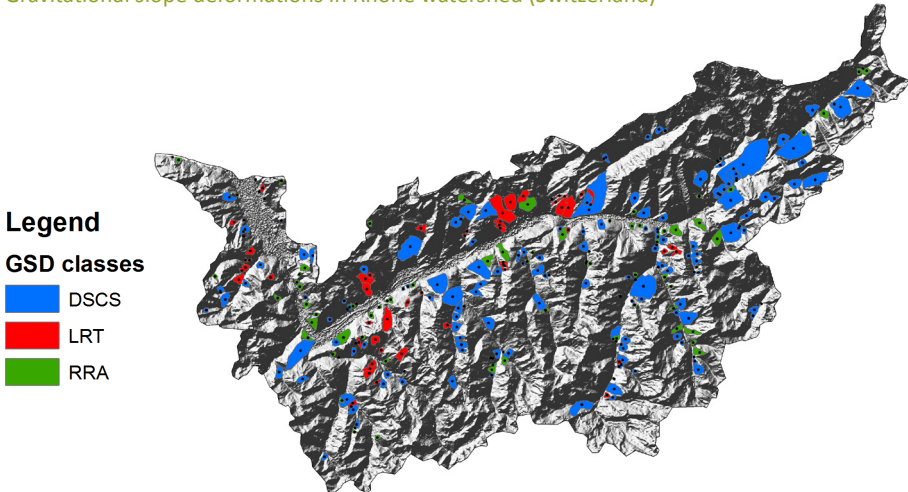
Case studies

1. The use of K-function to detect spatial pattern of landslides

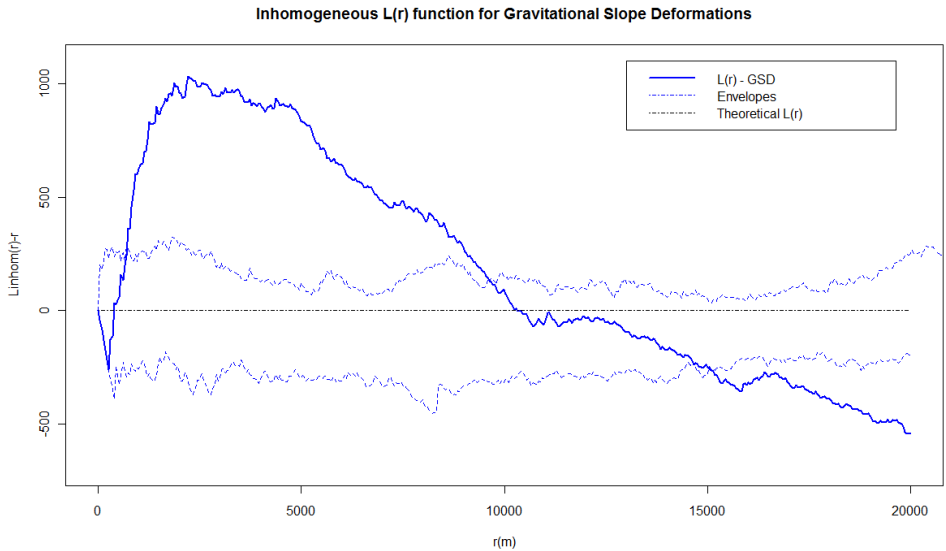
In the present study the spatial pattern of mapped landslides inventoried in the Rhone watershed (Switzerland) was analysed. A total of 294 gravitational slope deformations (GSD) were detected and classified based on their typology. The complete GSD geo-database was implemented at the Institute of Geomatics and Analysis of Risk (Lausanne University, Switzerland). The identification of the landslides was based on different sources of information such as geological maps, aerial photos and orthophotos, digital elevation model. The inventoried GSDs were furthermore classified following the modified version of the Hutchinson (1988) [9], resulting in three main classes: Rockslide and Rock-Avalanches (RRA), Deep Seated Creep/Sagging (DSCS) and Large Roto-Translation slides (LRT). The main objective of the present study is to verify if these events show a clustered or random distribution, and if it exists a spatial attraction (i.e. cluster) at a specific distance among the different classes of landslides.

FIGURE 1

Gravitational slope deformations in Rhone watershed (Switzerland)



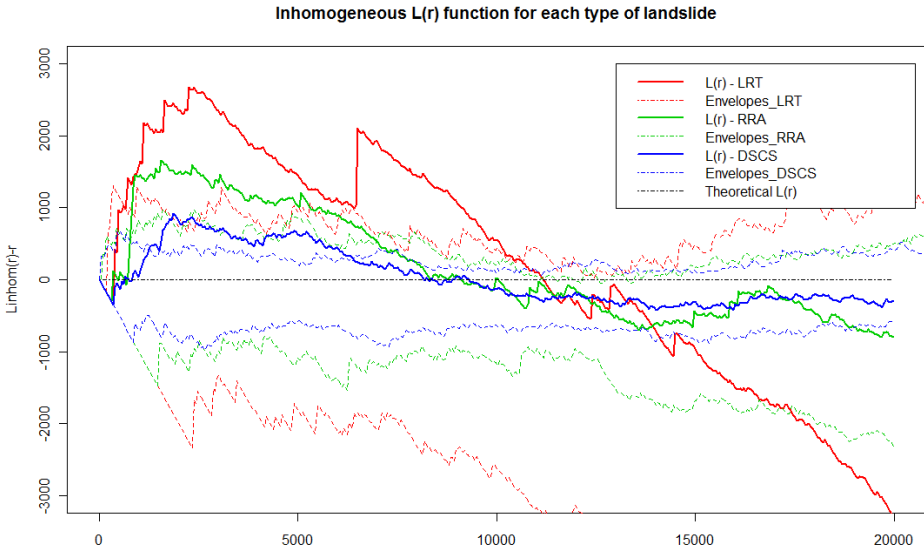
To test for randomness, the Ripley's K-function [15] was computed: the difference among the K-functions calculated for each dataset helped to compare the individual spatial pattern of each type of landslide and to reveal if they display similar cluster behaviour. Analytically $\lambda K(r)$ (where λ is the intensity of the point process) equals the expected number of additional points within a distance r from a randomly distributed event. Under complete spatial randomness (CSR) the theoretical $K(r)$ is equal to πr^2 . The estimated $K(r)$ can be plotted against the distance r and compared with the theoretical one: if the estimated function at a given distance r is higher than πr^2 , events are spatially clustered, whilst smaller values indicate repulsion between events. That way allows finding out at which range of distance data perform a non-random pattern distribution. To account for the natural non-uniform distribution of the geological events along the study area, a generalisation of $K(r)$ for a spatial inhomogeneous distribution was applied [2]. Edge correction was also introduced in the computation.

FIGURE 2Inhomogeneous $L(r)$ function for Gravitational Slope Deformations

In the present study we used a transformation of the Ripley's K-function, namely the L-function [5] that makes easier to compare the estimated with the theoretical curve and to evaluate departures from this last one. $L(r)$ equals the square root of $K(r)$ over π , minus r , so that the $L(r)$ theoretical value is zero at every distance. To test for spatial randomness, 999 Monte Carlo simulations of a realisation of an inhomogeneous random point process were performed to provide confidence envelopes. Results show that GSDs are not randomly distributed over the study area: indeed they are clustered at a distance ranging from about 75 m up to about 10 km, dispersed above about 15 km and included between the upper and the lower simulated curves in between, meaning a random distribution in this range.

FIGURE 3

Inhomogeneous $L(r)$ function for for each type of landslide



The cluster behaviour shows a maximum at about 2.5 km: this value can be retained for future local cluster analyses aiming to locate clusters in space. The $L(r)$ -functions computed individually over the single datasets show that the three classes of landslides have similar pattern behaviour, with a cluster tendency at a distance ranging from 500 m up to about 10 km, and that events belonging to LRT are more clustered than the events belonging to the RRA and DSCS landslide types.

2. Landslide recognition in LiDAR point clouds using NN-clutter removal

Light Detection and Ranging (LiDAR) is a remote sensing technique that allows obtaining the geometry of the terrain through the detection of the distance from the sensor to a given target. When mounted over a ground-based sensor, the so called Terrestrial Laser Scanner (TLS) is able to acquire

FIGURE 4

LiDAR points cloud

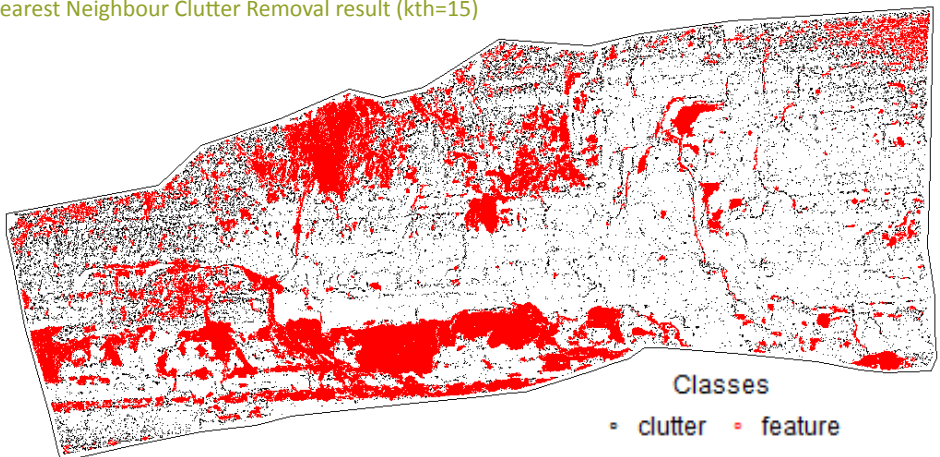


great resolution 3D information of the terrain (e.g. over 100 point per square meter) which results in a more or less homogeneously distributed point cloud. These points can be used to extract features' information of vegetation canopy, rivers, canyons, etc. Nevertheless, the automatic extraction of these features is not an evident task and different methods are currently being developed in different domains [8,17,10]. In the present study, we analysed TLS point clouds in order to automatically detect and extract landslides occurred in a pilot study area (Puigcercos, Catalonia, Spain).

The study area corresponds to a cliff affected by a high number of rockfalls per year [1]. The almost vertical geometry of the cliff allowed georeferencing points based on X,Z coordinates and to adopt a 2D approach.

FIGURE 5

Nearest Neighbour Clutter Removal result (kth=15)



The method is based on the Nearest Neighbour Clutter Removal (NNCR) in combination with the Expectation–Maximization (EM) algorithm [6]. The NNCR algorithm consists in compute the distance to the k th nearest neighbour for each point in the pattern: intuitively the points inside regions of higher density (i.e. in the features) have a smaller k th distance than the points inside regions of lower density (i.e. in the clutter).

Then the EM algorithm fits a mixture distribution (the feature and the clutter) to the nearest neighbour distances, which is used to classify each point as belong to the class “feature” or “clutter”. The degree on neighbour (k) has to be fixed from the user and this choice can affect the result of the analysis. In the present study several increasing values for k were applied and then the one based on an entropy type measure of separation was retained.

This method was applied after the pre-filtering of the LiDAR point cloud in order to remove points affected by instrumental error. The density of points belonging to landslides is higher than the density of points falling outside and the applied technique allowed assigning them to the class “feature” with their estimated probabilities. This preliminary result can be used to calculate landslide volume or to analyse landslide spatio/temporal patterns. The proposed method allowed automatically detecting landslides gaining in times and in accuracy compared to the visual technique.

Further development of the research will consist in the custom implementation of the algorithms in the R environment to be able to analyse multi-dimensional point patterns, including the third spatial dimension and/or the temporal dimension.

[1] ABELLÁN, A., CALVET, J., VILAPLANA, J.M., AND BLANCHARD, J. Detection and spatial prediction of rockfalls by means of terrestrial laser scanner monitoring. *Geomorphology*, 119 (2010), pp. 162-171.

- [2] BADDELEY, A., MOLLER, J., WAAGEPETERSEN, R. Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54, (2000), pp. 329–350.
- [3] BADDELEY A., AND TURNER R. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12 (6) (2005), pp. 1–42.
- [4] BAI, S.B., WANG, J., LÜ, G., ZHOU, P., HOU, S.S., XU, S.N. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the *Three Gorges area, China*. *Geomorphology*, 115 (2010), pp. 23–31.
- [5] BESAG, J. Discussion of Dr Ripley's paper. *Journal of the Royal Statistical Society, Series B*, 39, (1977), pp. 193–195.
- [6] BYERS, S., AND RAFTERY, A. E. Nearest-Neighbor Clutter Removal for Estimating Features in Spatial Point Processes. *Journal of the American Statistical Association*, 93, (1988), pp. 577–584.
- [7] ERENER, A., AND DÜZGÜN, H.S.B. Landslide susceptibility assessment: what are the effects of mapping unit and mapping method? *Environmental Earth Sciences*, June 2012, Volume 66, Issue 3, pp. 859–877.
- [8] GIGLI, G, AND CASAGLI N. Semi-automatic extraction of rock mass structural data from high resolution LiDAR point clouds. *International Journal of Rock Mechanics and Mining Sciences*, Volume 48, Issue 2, February 2011, pp. 187–198.
- [9] HUTCHINSON, J. N. General Report: Morphological and geotechnical parameters of landslides in relation to geology and hydrogeology. *Proceedings of the Fifth International Symposium on Landslides*. Edited by: Bonnard, C., Balkema, Rotterdam, pp.3–35 (1988).
- [10] KABOLIZADE, M., EBADI, H., AHMADI, S. An improved snake model for automatic extraction of buildings from urban aerial images and LiDAR data. *Computers, Environment and Urban Systems*, Volume 34, Issue 5, August 2010, pp. 435–441.
- [11] LEE, S., RYU, J.-H., KIM, I.-S. Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: Case study of Youngin, Korea. *Landslides*, 4 (2007), pp. 327–338.
- [12] NANDI, A., AND SHAKOOR A. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*, 110 (2010), pp. 11–20.

- [13] OH, H.J., AND LEE, S. Landslide susceptibility mapping on Panaon Island, Philippines using a geographic information system. *Environmental Earth Sciences*, 62 (2011), pp. 935–951.
- [14] R DEVELOPMENT CORE TEAM (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved September 7, 2012 from <http://www.R-project.org/>
- [15] RIPLEY, B.D. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B*, 39, (1977), pp. 172-212.
- [16] VAN DEN EECKHAUT, M., POESEN, J., GULLENTOPS, F., VANDEKERCKHOVE, L., HERVÁS J. Regional mapping and characterisation of old landslides in hilly regions using LiDAR-based imagery in Southern Flanders. *Quaternary Research*, Volume 75, Issue 3, May 2011, pp. 721-733.
- [17] ZHAO, K., POPESCU, S., MENG, X., PANG, Y., AGCA, M. Characterizing forest canopy structure with LiDAR composite metrics and machine learning. *Remote Sensing of Environment*, Volume 115, Issue 8, 15, August 2011, pp. 1978-1996.
- [18] ZUO, R., AGTERBERG, F.P., CHENG, Q., YAO, L. Fractal characterization of the spatial distribution of geological point processes. *International Journal of Applied Earth Observation and Geoinformation*, Volume 11, Issue 6, December 2009, pp. 394-402.

Geographical analysis and numerical quantification of visual impact for aerogenerators and photovoltaic panels using Open Source GIS

Authors

- **Annalisa Minelli**, GfosServices S.A. (www.gfoservices.it); Dipartimento di Ingegneria Civile e Ambientale, Università degli Studi di Perugia, *Italy*
- **Ivan Marchesini**, GfosServices S.A. (www.gfoservices.it); CNR IRPI, Perugia, Geomorphology Research Group, Italy; T4E S.r.l. (www.t4e.it) *Italy*
- **Pierluigi De Rosa**, GfosServices S.A. (www.gfoservices.it); Dipartimento di Ingegneria Civile e Ambientale, Università degli Studi di Perugia, *Italy*
- **Luca Casagrande**, GfosServices S.A. (www.gfoservices.it); T4E S.r.l. (www.t4e.it), *Italy*
- **Michele Cenci**, Regione Umbria, sezione «Cave, miniere, rischio di incidente rilevante», *Italy*

KEYWORDS : visual impact, photovoltaic panels, aerogenerators, landscape, GIS

Landscape is not only a static fact, but its quality can influence the quality of life of each citizen, since the environmental psychology [9] and the geopsycology can affect the each one's choices and preferences [3]. Photovoltaic fields and windfarm can really impact on the landscape perception.

The background of this work is represented by some optical studies about the human eye perception and the definition of the field of view (for example

[12], [2]) some studies about descriptive geometry [4] and Geographical Informative Systems competencies. In particular the computation of the right field of view for a specific distance and position between object and observer, the evaluation of distortion perceived by the human eye, while observing the object and the capability to analyze the mutual relation between object, observer and earth morphology, allows to evaluate quantitatively the real visual impact of this kind of systems on the landscape.

The visual impact of photovoltaic fields and windfarm is one of the most tricky themes in landscape quality evaluation because there is still not a specific, univocal and standardized method to quantify the impact and there are not guidelines internationally adopted for geographical “non qualitative” analyses, but only qualitative or state-specific laws (for example [10], [5], [6]).

This note shows a geographical method to evaluate quantitatively the visual impact of aerogenerators and photovoltaic panels. The software used are two Open Source GIS: GRASS GIS [7] and QGIS, and the programming language Python. In particular, a specific tool called “r.wind.sun” has been developed using the modules of GRASS GIS and the Python scripting and then the tool has been added to the GRASS GIS plugin tool box of QGIS. The adoption of QGIS was needed in order to facilitate the use of the module from the public administrators who can find the GRASS GIS graphical user interface (GUI) a little tricky. A great enhancement for the module has been the changeover, in the use of the visibility tools, from r.los (GRASS GIS version < 6.4) to r.viewsheed (GRASS GIS version 7.0), which we have sponsored.

The rationale of the module is that the visual impact of an object is related to its occupancy of the human field of view. In order to quantify this impact, the portion of the areas belonging to the observed object, respect to the human field of view (FOV), is calculated. In this way, the visual impact is represented, in the end, by a nondimensional intervisibility index (NI-ratio). In particular, taken as fixed the dimension of the observed object and the angles (up, down, left, right) defining the human field of view, the FOV area only depends on the distance from the observer. The NI-ratio is function of the FOV area calculated for the distance existing between the observer and the object.

The placement of the aerogenerator or of the photovoltaic panel is generally known and that means that the NI-ratio can be calculated for each point in the surroundings. In order to run the module user needs a digital elevation model (DEM) and a vector point layer containing the position of the aerogenerators or of the photovoltaic panel centroids. Moreover some other specific information (aerogenerators dimensions, photovoltaic panel dimensions and inclinations) have to be provided. The output of the model is a raster layer where cell values represent the nondimensional intervisibility index value. Optionally, for the photovoltaic panels, a layer of areas eventually prone to dazzling (due to panel's surface reflection) can be generated.

The analysis is executed singularly for each element (single panel or aerogenerator), in the way that, for each one of them, a raster map of the non dimensional impact index is created. Once every panel or aerogenerator is processed, the final map is obtained by sum the impact index value for each map, on each pixel of the final map. Moreover, in estimating the area of each object, both the distance between the object and the observer and the inclination of line which links them, are considered; in the way that the object results with a certain distortion (the panel surface, which is rectangular becomes irregular, the rotor of the aerogenerator occupies, with its rotation, an ellipsoidal area), as it is really perceived by the human eye.

Since the map produced by the model are in raster format, it is understandable that the duration of each simulation is strongly related to the maximum distance chosen to evaluate visual impact. Moreover, the values obtained for the impact index are often very small. This happens because when the distance between the object and the observer grows, the human field of view increase more than the object gets small, so, the values of impact index obtained by the simulation, decrease in lognormal way with the distance between observer and object. To better understand the results of each simulation, it can be useful to reclassify the output map, with the aim to obtain a zone-map from "small" to "great" impact (Figure 1). Another way to better understand a resulting value (NI-ratio) is to estimate the distance from the observer at which, a little object of known dimensions (for example a paper sheet), have to be put in order to give the same occupancy of field of view (NI-ratio).

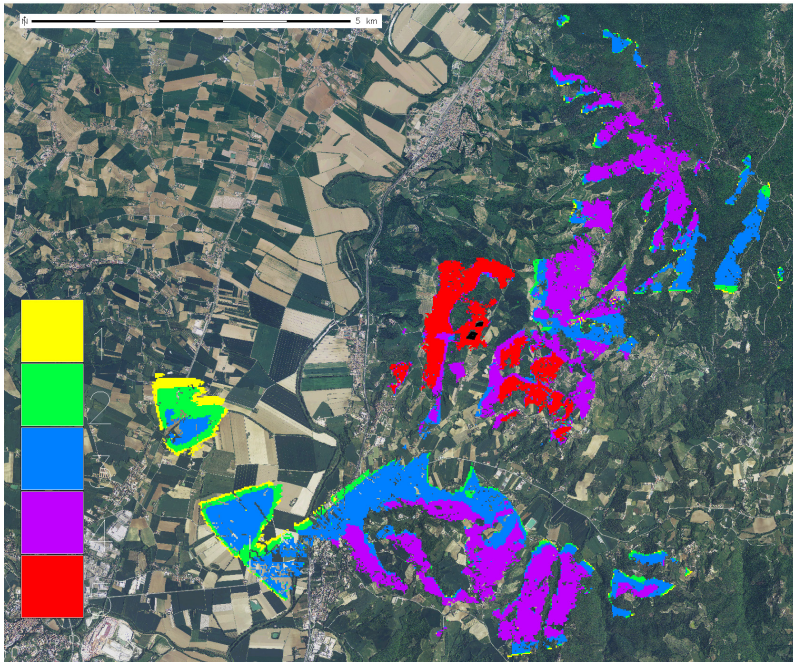


FIGURE 1

To simplify the use of this kind of tool and to keep the same scientific power, we are planning to serve this model as a Web Service using the Web Processing Service (WPS) standard from OGC [8]. With this approach the tool could be executed using a WebGIS application with a user friendly interface and with no need to install any other tool than a Web Browser. In this way it is possible for each one who wants to evaluate the impact, to use this methodology with just few clicks and, for example using a Mobile Device, to run the tool directly in place and to see the result using Augmented Reality [1]. Especially in mobile application, to grant compatibility, is very important to develop tool which can be executed from a Web Browser.

Since there is not a precise set of rules in this specific field, but only some similar studies involving social or agricultural factors (for example [11]), the method allows to evaluate the intervisibility in a purely objective way by providing not a judgement, but a representative number. Moreover, this method can be useful to professionals and public administration to decide the feasibility of the photovoltaic field or the windfarm.

- [1] BIMBER, O., RASKAR, R. *Spatial Augmented Reality - Merging Real and Virtual Worlds*. A K Peters Ltd (2005); ISBN: 1-56881-230-2.
- [2] COSTELLA, J. P. *Galilean Antialiasing for Virtual Reality Displays*. sci.virtual-worlds,4 November 1992; University of Melbourne preprint UM-P-92/105.
- [3] CRAIK, K. H. Multiple scientific paradigms in Environmental Psychology. *International Journal of Psychology*, 12, 147- 157 (1977).
- [4] DE RUBERTIS, R. *Fondamenti di geometria descrittiva*. Edizioni Kappa, (1979) Roma.
- [5] HURTADO J.P., FERNANDEZ J., PARRONDO J.L., BLANCO E. Spanish method of visual impact evaluation in wind farms. *Renew Sustain Energy Rev* 2003; 8:483–91.
- [6] LADENBURG, J. Visual impact assessment of offshore wind farms and prior experience. *Applied Energy*, 86 (2009), pp. 380–387.
- [7] NETELER, M., MITASOVA, H. *Open Source GIS: A GRASS GIS Approach*. Third Edition. The International Series in Engineering and Computer Science: Volume 773. 406 pages, 80 illus., Springer, New York; ISBN: 038735767X (2008).
- [8] OPEN GEOSPATIAL CONSORTIUM INC *OpenGIS Web Processing Service*. OGC Publicly Available Standard. Reference number: OGC 05-007r7. Editor: Peter Schut (2007).
- [9] POL, E. Blueprints for a History of Environmental Psychology (I): From First Birth to American Transition. *Medio Ambiente y Comportamiento Humano*, 2006, 7(2), 95-113. ISSN 1576-6462.
- [10] REGIONE AUTONOMA DELLA SARDEGNA *Linee guida per l'individuazione degli impatti potenziali degli impianti fotovoltaici e loro corretto inserimento nel territorio [Guidelines for the assessment of the potential impacts of photovoltaic systems and for the correct siting]*. Decree n. 30/2 on the 23 May 2008.
- [11] ROGGE, E., NEVENS, F., GULINK, H. Reducing the visual impact of 'greenhouse parks' in rural landscapes. *Landscape and Urban Planning* 87 (2008) 76–83.
- [12] SPECTOR, R. H. *Visual Fields. Clinical Methods: The History, Physical, and Laboratory Examinations*. 3rd edition. Boston: Butterworths; 1990. Chapter 116.

Relatedness and scale dependency in very high resolution digital elevation models derivatives

Authors

- **Kevin Leempoel**, EPFL, *Switzerland*
- **Stéphane Joost**, Ecole Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), GIS Research Laboratory (LASIG), *Switzerland*

KEYWORDS : Very High Resolution (VHR), Digital Elevation Model (DEM), Environmental variables, Multiscale, SAGA GIS, Rochers-de-Naye (Swiss prealps)

Introduction

In the field of landscape genetics, environmental variables are needed to assess the influence of the environment on the spatial distribution of the genetic information of animals or plants [1]. However, measuring variables (e.g. temperature, precipitation) at locations of each sampled individuals is time consuming. Landscape geneticists therefore often use variables from remote sensors such as existing satellite imagery and digital elevation models (DEMs). From these DEMs, many variables related to morphology, solar radiation or hydrology can be computed. However, their spatial resolution is not high enough when studying the local adaptation of organisms to their local environment. Indeed climatic variables available for our study site in Switzerland offer a resolution that is coarser than 25m [2] and the same issue is encountered for the globally available DEMs which have a spatial resolution of 30 (ASTER) and 90m (SRTM) with a low accuracy (+- 15m) [3, 4]. An additional difficulty comes as many indices have been created in the last 20 years to approximate wetness, erosion, soil type distribution, stream power, etc. These indices are often highly mutually correlated and the impact

of spatial resolution on their computation is unknown. Also, the user faces the choice of the software, which are numerous and are able to provide different variables computed with different methods.

In this paper we use the software SAGA GIS, which has the advantages to be both free and open source, but remains rarely used in landscape genetics. SAGA is a raster GIS and is able to produce variables that are not proposed by other software, making it one of the most complete to analyze DEMs. These variables are described in papers informing on the context and their relatedness to environmental processes. In addition, SAGA GIS can be accessed through R. It is thus possible to script once the computation of all variables for a model instead of computing them “by hand” in the software. We computed 17 environmental variables (see Table 1) using SAGA software for three DEMs of different spatial resolution (0.25, 0.5, 2m) acquired over the region of “les Rochers de Naye” where we study the adaptation of the alpine plant *Biscutella Laevigata* L. (Brassicaceae) to its environment. We compared these 51 variables in order to: I) assess their mutual correlation, II) analyze the influence of the spatial resolution on their computation, III) assess whether a DEM obtained from stereophotogrammetry is accurate enough to compute these variables.

Material and Methods

Study site

“Les Rochers-de-Naye” study area is located in the Swiss Prealps (N46°26′00“, E6°58′50“). Altitude ranges between 1805 and 2030m. The site is characterized by a steep South-East-facing ridge covered mostly by grass. We showed in a recent paper that the use of variables related to hydrology, topography and solar radiation computed from DEMs is relevant to identify genomic regions possibly involved in the adaptation process of *Biscutella laevigata* to its local environment [5].

Digital Elevation Models

SwissAlti3D (ST2m) is a 2m spatial resolution model acquired by swisstopo using a LIDAR sensor mounted on a plane. Such VHR model is rarely available in other countries. Elevation accuracy is expected to be $\pm 0.5\text{m}$ up to 2'000

meters. (<http://www.swisstopo.admin.ch/internet/swisstopo/en/home/products/height/swissALTI3D.html>).

Rpod0.5m was acquired by the Rpod team using a drone and the SenseFly technology. Images were used to produce a DEM by means of stereophotogrammetry (<http://www.r-pod.ch/>; <http://www.sensefly.com/>). The spatial resolution is 0.5m and the accurateness in elevation is often lower than a LIDAR based model.

Finally, Heli0.25m was acquired by Helimap using a LIDAR sensor mounted on a helicopter (<http://www.helimap.ch/>). After filtering the laser point cloud, the spatial resolution was set to 0.25m.

Derivated variables

In SAGA GIS, we computed the 17 variables shown in table 1 for each of the 3 DEMs. Some of these are of interest for ecology and landscape genetics studies. The dataset used to calculate the correlations is defined by 1000 points randomly distributed over the study area. Values of all variables were extracted for each of these points using the nearest neighbor method.

Hypothesis of independence between variables as well as correlations coefficients were calculated in R [12] for each pair of variable. Only significant correlations ($p < 0.01$ for a Spearman's rank correlation test) are discussed. In addition, for each trio of a variable (i.e. the same variable computed for each DEM), box percentile plots were created in R and correlation coefficients compared.

Results and discussion

Curvature, slope length and catchment area show few significant correlations, unlike solar radiation variables that are correlated to almost all variables. In fact, excepted for Direct-to-diffuse ratio, all variables related to insolation are mutually highly correlated (0.70 – 0.97). Consequently, considering only one of the three variables (i.e. Direct insolation, Duration of insolation or Total insolation) for one period during the year is enough to approximate a relative

insolation at our study site. However, for direct-to-diffuse ratio computed for December 21 and June 21, the correlations with other insolation variables can be low (several ≈ 0.5). While this may be meaningful for plants, it is difficult to analyze the pertinence of such variable since diffuse insolation is computed on the basis of a uniform reflection of the surface, which may be inaccurate as, for example, snow is present until May on our study site. Another example of redundancy exists between MPI and sky view factor, with a strong negative correlation at all scales ($-0.79 - -0.93$). In addition correlation coefficients are similar between the different spatial resolutions for sky view factor ($0.82 - 0.92$) and for MPI ($0.83 - 0.94$). Consequently, the computation of one of these variables from one DEM is sufficient to represent the heterogeneity of the landscape.

High correlation values are calculated between the same variables at different resolutions. First, as expected, the correlation between altitude models is very high (>0.99); however this does not imply a strong correlation between all variables calculated from these DEMs, as explained for example in [13]. For the first order derivatives, aspect and slope show high correlations between resolutions ($0.90 - 0.96$; $0.84 - 0.90$ respectively) while second order derivatives (Plan and profile curvature) vary a lot with resolution (no significant correlation). This observation on simple terrain attributes is coherent with the literature over other areas [13, 14].

MPI correlation between different resolutions is high as mentioned above. As this index is based on the highest and lowest angle found over a defined radius, it depends more on the altitude precision than on the resolution of the DEM. However, an interesting observation is that the range of values for this index is much wider in the Rpod0.5m than in the two LIDAR models. This remark is also valid for the TRI despite showing more variation among the different spatial resolutions ($0.61 - 0.75$). The most likely explanation of the higher correlation between LIDAR DEMs than between LIDAR DEMs and Rpod DEM would be due to the technique used (stereophotogrammetry) and to the filter applied. In fact, the additional noise due to its lower accuracy influences these indices that depend on the heterogeneity of the landscape.

TABLE 1

	Variable	Unit
<u>Variables related to morphology</u>	Slope	0-90 Degrees
	Aspect	0 (Nord) -360 degrees
	Plan Curvature	°/100LU
	Profile Curvature	°/100LU
	Terrain Ruggedness Index (TRI)	
	Morphometric Protection Index (MPI)	no unit 0-1
	Sky view factor	no unit 0-1
<u>Solar radiation variables</u>	Direct insolation	kwh/m ²
	Total insolation	kwh/m ²
	Direct-to-diffuse ratio	no unit
	Duration of insolation	hours
<u>Variables related to hydrology</u>	Modified catchment area	m ²
	Catchment slope	degrees
	Slope length	m
	Stream power index	no unit
	LS factor	no unit
	Wetness index	no unit

Parameters	Comment	Reference
Method= Fit 2.Degree polynom		[6]
	LU is the length unit	
	LU is the length unit	
Radius = 2meters	Quantifies the heterogeneity of the topography	[7]
	Degree of dominance or enclosure of a point on an irregular surface. Evaluates the protection of a point from the surrounding relief over a defined distance	[8]
Distance = 20 meters	Ratio of the sky area over the obstructed area	[9]
Parameters: Atmospheric effects=lumped atmospheric transmittance, Latitude = 46°, time resolution = 0.5(h); period= 21 December and 21 June		[9]
	Discharge contributing upslope area of each grid cell	[10]
	Average slope over the catchment	[10]
	Mean length of flow paths to a point in the catchment	
	$SPI = 0.2 (Specific\ catchment\ area \times plan\ curvature \times \tan(slope))^{0.25}$	[11]
	$LS = 1.4 * \left(\frac{Catchment\ Area}{22.13}\right)^{0.4} \times \left(\frac{\sin(slope)}{0.0896}\right)^{1.3}$ effect of slope length on erosion	[11]
	$WI = \ln\left(\frac{Specific\ catchment\ Area}{Slope}\right)$	[10]

Hydrology and insolation variables were both expected to show an effect of fine scale heterogeneity. This was not observed for the latter ones and it could be explained by the terrain attributes they depend on. Indeed, insolation variables depend on orientation, slope and shadow of surrounding relief, which are not varying much with resolution, as mentioned above. Quite the reverse, hydrology variables are based on curvature, for which scale variability is high. The fine scale structure of Rpod0.5m and Heli0.25m may allow representing these small variations that alter the local flow and the consequence being that SPI, WI, LS factor, slope length and Modified catchment area have smaller values on average for these two DEMs. As regards catchment slope, this variable does not vary as it is based on the average upslope.

Our results also demonstrate that the Rpod DEM is showing similar results like the ST2m and the Heli025m for most variables while being much less expensive and easier to acquire. We show with these high correlations that the method based on stereophotogrammetry, which is usually less accurate, should be considered in the landscape genetics domain to produce new high resolution environmental variables. Indeed, for most regions in the world, freely available DEMs have a low resolution and a low accuracy (ASTER and SRTM model mentioned in introduction). Therefore acquiring a DEM using a drone can be of great interest for research purposes in order to acquire variables at a fine scale.

[1] MANEL, S., SCHWARTZ, M. K., LUIKART, G. AND TABERLET, P. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18, 4 (2003), 189-197.

[2] ZIMMERMANN, N. E. AND KIENAST, F. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *J Veg Sci*, 10, 4 (1999), 469-482.

[3] TACHIKAWA, T., HATO, M., KAKU, M. AND IWASAKI, A. Characteristics of Aster Gdem Version 2. *Int Geosci Remote Se*(2011), 3657-3660.

- [4] FARR, T. G., ROSEN, P. A., CARO, E., CRIPPEN, R., DUREN, R., HENSLEY, S., KOBRICK, M., PALLER, M., RODRIGUEZ, E., ROTH, L., SEAL, D., SHAFFER, S., SHIMADA, J., UMLAND, J., WERNER, M., OSKIN, M., BURBANK, D. AND ALSDORF, D. The Shuttle Radar Topography Mission. *Rev. Geophys.*, 45, 2 (2007).
- [5] LEEMPOEL, K., STUCKI, S., PARISOD, C. AND JOOST, S. Very high resolution digital elevation models (VHR DEMs) and multiscale landscape genomics analysis applied to an alpine plant species. In *Proceedings of the SIGSPATIAL Special (2011)*, ACM press, 10-14.
- [6] ZEVENBERGEN, L. W. AND THORNE, C. R. Quantitative-Analysis of Land Surface-Topography. *Earth Surf Processes*, 12, 1 (1987), 47-56.
- [7] RILEY, S. J., DEGLORIA, S. D. AND ELLIOT, R. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences*, 5, 1-4 (1999), 23-27.
- [8] YOKOYAMA, R., SHIRASAWA, R. J. AND PIKE, R. J. Visualizing topography by openness: A new application of image processing to digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 68 (2002), 251-266.
- [9] BÖHNER, J. AND ANTONIĆ, O. Chapter 8 Land-Surface Parameters Specific to Topo-Climatology. Elsevier, City, 2009.
- [10] BÖHNER, J., KÖTHE, R., CONRAD, O., GROSS, J., RINGELER, A. AND SELIGE, T. Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation, EUR 20398 EN(2002).
- [11] MOORE, I. D., GRAYSON, R. B. AND LADSON, A. R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol Process*, 5, 1 (1991), 3-30.
- [12] R DEVELOPMENT CORE TEAM R : A language and environment for statistical computing. R Foundation for Statistical Computing, City, 2010.
- [13] WILSON, J. P. AND GALLANT, J. C. *Terrain Analysis : Principles and applications*, New York, 2000.
- [14] LASSUEUR, T., JOOST, S. AND RANDIN, C. F. Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological Modelling*, 198, 1-2 (2006), 139-153.

Open source software for Big Data : Experiences in indexing and browsing geo-archival records

Authors

- **Jefferson Robert Heard**, RENCI, UNC Chapel Hill, *United States*
- **Richard Marciano**, UNC Chapel Hill, *United States*

KEYWORDS : H.3.1 Content Analysis and Indexing, I.7.5 Document Analysis, I.7.3 Index Generation, H.3.7 Digital Libraries, H.2.8 Spatial Databases and GIS

Introduction and Related Work

Archival records are not like records in a database. The original structure, attributes, and metadata of an archival record are as important to archival integrity as the record itself. These attributes include things like file structure, directory hierarchy, and file permissions. Organic, human structures do not always match up well with storage and retrieval patterns. Thus an archive of hundreds of millions or a billion records becomes a big data problem.

Additionally, certain kinds of records are difficult to browse and search. Much work has been done in making interfaces for searching text. Less work has been done on making geographic records searchable or browsable. In this talk, we present open source approaches to indexing and browsing geographic records in a large archival collection.

Some amount of work has been done in browsing collection of archival data. In particular, treemaps have been employed [1] along with metadata visualization to enable browsing of archival metadata and records. Large amounts of work have been done on how to effectively store archival data. In particular, IRODS [2], and Globus Online [4] have emerged as “Data Grid” software that provide rich capabilities to a user interested in preserving a dataset.

In this presentation we will detail our experiences with developing a system to scalably index and browse geographic records in an archival setting.

Approach

Our archival collection is known as the CI-BER Testbed. It is a still-growing 70 million file, 41TB collection consisting of data from across 133 different agencies of the US government. The CI-BER Testbed was developed to create a system for testing the scalability of archival systems across many kinds of heterogenous data, from files consisting of gigantic chunks to directories containing hundreds of thousands of files, to deeply nested structures. The CI-BER Testbed is housed on an iRODS data grid that has been federated with the US National Archives and Records Administration (NARA)'s own data grid.

IRODS, the Integrated Rule Oriented Data System is an open source data grid software originally developed by the DICE Group. It provides a file-system abstraction that allows for arbitrary file, directory, and file-system level metadata, a rich permissions system, user-defined “rules” that govern how data is collected into the filesystem, and “microservices” that can operate as independent agents on a collection. IRODS is used extensively in the academic and archiving world to manage large data collections.

We have built a distributed indexing system that authenticates with this data grid and crawls individual collections looking for geographic files. An administrator sends a web request to index a particular collection, then the indexer uses the user's IRODS credentials to mount the IRODS collection and crawls that collection for filenames. The filenames are then farmed out as individual tasks to each of the worker threads. Each worker thread attempts first to open the file with GDAL and then OGR [3]. If either of these works, then the worker thread passes the opened file to a number of metadata extraction functions and the metadata, along with the file, is added to the index.

For the index, we selected an open-source “NoSQL” database, MongoDB [5]. MongoDB was selected both for its scalability to large amounts of data and its

schema-free nature. We wanted the index to be extensible, so that alternative methods for metadata extraction could continue to append metadata to each individual record. These metadata appendices to each record could then be visualized independently for the records that have them. Core to our metadata, however, are the geographic boundaries of a particular dataset, translated into WGS84, the name of the dataset, and for coverages, the number of bands, or for feature sets, the name of layers and for each layer the name and type of record fields.

Directory paths within the archives have metadata from the individual records aggregated and appended. This results in each directory level having a boundary box associated with it as well as the number of archival records contained in that directory.



FIGURE 1
The indexer

This MongoDB index is used to power our visualizations. Our main visualization is a browsing interface for the records based on the open source JQueryMobile [6], D3 [7], and OpenLayers [8]. In the visualization, we extend the treemap for use specifically with geographic records by adding a geographic map to the interface and unifying the two.

The user is initially presented with an overview. The overview centers the map on the boundary of the entire collection set. The first level of the treemap contains one cell for each indexed collection. Treemap cells are sized by the number of actual record items contained within the collection. For levels containing leaf nodes in the treemap, each cell is sized by the physical area of the bounding box (in square meters) relative to its peers. Each collection has its own individual bounding box on the map. If the user touches a bounding box, it highlights the corresponding object in the treemap. If the user touches the treemap, the collection is descended into and the app retrieves the next level of the index.

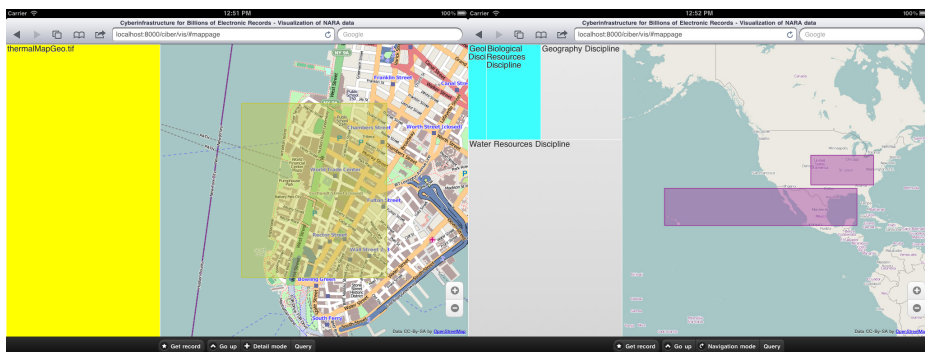


FIGURE 2

Detail and overview showing the treemap's grouping of records and their location on the map

Additionally, there is a “detail mode.” If the user touches on a treemap cell the app will center and expand on the cell's bounding box. Tapping “get record” retrieves the metadata record associated with the bounding box. For an individual record, this is the geographic metadata associated with the file. For a subcollection, this is the bounding box and the number of records in the subcollection. That metadata record is visualized and a link to the record itself is provided.

Future Directions

In the near future, we intend to incorporate this visual browsing interface into larger, interactive mapping applications. These applications will serve to encourage community involvement in archives and a view of archives as a first-class Open Source medium.

Acknowledgements

This work was sponsored by OCI-084296, a grant under the US National Archives and Records Administration (NARA) Applied Research division and the US National Science Foundation Office of Cyberinfrastructure (NSF/OCI)

[1] ESTEVA, M., XU, W., JAIN, S.D., LEE, J.L., MARTIN, W.K., Assessing the preservation condition of large and heterogeneous electronic records collections with visualization, *International Journal of Digital Curation* 6(1):45-57 Feb. 2011.

[2] ZHU, B., MARCIANO, R., MOORE, R., HERR, L., SCHULTZ, J., Digital Repository: Preservation Environment and Policy Implementation, Springer Verlag 2012, *International Journal on Digital Libraries (IJDL)*.

[3] THE OSGEO GROUP. GDAL and OGR, 2012. <http://www.gdal.org/>

[4] THE GLOBUS GROUP. *Globus Online*, 2012. <http://www.globusonline.org/>

[5] 10GEN, INC. *MongoDB*, 2012. <http://www.mongodb.org/>

[6] THE JQUERY GROUP. *JQueryMobile*, 2012. <http://jquerymobile.org/>

[7] BOSTOCK, M. D3.js. *Data Driven Documents*, 2012. <http://www.d3js.org/>

[8] OPENGEO, ENC. *OpenLayers*, 2012. <http://www.openlayers.org/>

Physical Landscape of Britain and Northern Ireland: technical development

Authors

- **Claudio Piccinini**, Kingston University, Faculty of Science, Engineering and Computing, Kingston Upon Thames, Surrey, London, KT1 2EE, *United Kingdom*
- **Mike Smith**, Kingston University, Faculty of Science, Engineering and Computing, Kingston Upon Thames, Surrey, London, KT1 2EE, *United Kingdom*
- **Janet Hooke**, University of Liverpool, Roxby Building, Liverpool, L69 7ZT, *United Kingdom*
- **Katherine Hesketh**, University of Liverpool, Roxby Building, Liverpool, L69 7ZT, *United Kingdom*

KEYWORDS : geomorphology, web mapping, spatial databases, PostGIS, Javascript, ExtJS, OpenLayers, OpenWebGlobe, PHP, Web storage

1. Background

Geomorphology is the science that analyses how climatic, tectonic and biogenic processes act on the surface of the Earth to create landforms and landscapes. Whilst many organizations recognize the importance of geomorphic processes and landforms in their work, there is no single centralised resource on the geomorphology of Britain. Some useful sources exist but these tend to be thematic (e.g. BRITICE), spatially restricted (e.g. GeoEast) or broad-brush (e.g. National Character Areas).

The aim of this project is, for the first time, to provide the interested professional, researcher and the general public with access to information, data and knowledge on the geomorphology of the British landscape.

Information will be accessible through an interactive web mapping application built using both open data (e.g. Ordnance Survey Open Data Gazetteer) and open source technologies (e.g. OpenLayers). The web application links to bibliographic details of academic research and reports enabling users to locate relevant research work for locations of interest. The web mapping application will be useful for the interpretation of landscape and acquiring the necessary knowledge for assessing environmental impacts, environmental change, environmental hazards and associated risks, and for a wider understanding of the physical environment.

The official web site for this project is available at www.landscapebritain.org.uk.

2. Database

To carry out the project three main components were developed: a database, a database front-end and a web mapping application. The database was designed for the storage and visualization of bibliographic references covering the geomorphology of the British landscape. Figure 1 shows the Entity-Relationship model which highlights the database structure and relationships between tables. The model is divided into 6 groups.

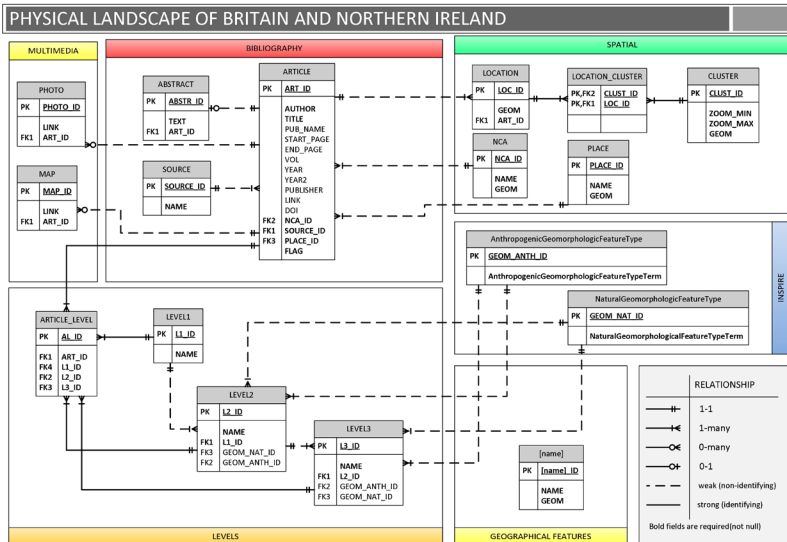


FIGURE 1

1. 'Bibliography' (red) contains the 'Article' table to store information about each reference. The abstract is stored in a separate table with a full-text search index; this will allow users to search for bibliographic records querying the abstracts with single words or phrases. Reference sources are classified as one of five types values: book, journal, dissertation, book chapter and unpublished report.
2. 'Multimedia' (yellow) stores images and static maps associated with the references that are visualized with the web mapping application.
3. 'Spatial' (green) holds information about the location of bibliographic records. The 'Location' table stores one or more polygonal outlines for each reference as a bounding box; this simplifies the database entry and increases the speed of database searches.

TABLE 1

LEVEL1	Environment/ Landform	Processes	Impact	Material	Timescale or period
LEVEL2	Aeolian	Accretion	Agricultural impact	Alluvium	Annual
	Bog	Advance	Climate change	Bedrock	Century
	Coastal	Denudation	Dam impact	Boulders	Decadal
	Downland	Deposition	Desertification	Clay	Devensian
	Escarpment	Erosion	Environmental impacts	Colluvium	Event
	Estuarine	Fire	Eutrophication	Consolidated	Historical
	Fells	Glaciation	Human impacts	Gravel	Holocene
	Fluvial	Groundwater	Land - use impact	Minerogenic	Longer
	Glacial	Hydraulics	Mining impact	Organic	Millenia
	Hills	Hydrology	Sea Level Rise	Peat	Pleistocene
	Karst	Isostasy	Storm surges	Sand	Quaternary
	Lacustrine	Mass movement	Tourism impact	Silt	Seasonal
	Loess	Migration	Urbanisation	Soil	Tertiary
	Lowland	Nutrient flux		Unconsolidated	
	Marine	Pollution			
	Moorland	Retreat			
	Mountain	Runoff			
	Periglacial	Sea-level			
	Plateau	Sediment transport			
	Slope	Sedimentation			
	Structural	Solutes			
	Upland	Storage			
	Urban	Storms			
	Wetland	Tides			
		Uplift			
		Water flow			
		Waves			
		Weathering			

TABLE 2

LEVEL2	Fluvial	Coastal	Mass movement	Slope	Aeolian
LEVEL3	Alluvial fan	Bar	Debris flow	Cliff (slopes)	Dune
	Bank	Beach	Landslide	Debris cones	Deposit
	Bar	Berm (coastal)	Mudflow	Debris flow	Process
	Braiding	Breach	Rockfall	Gullies	Landforms
	Buried channel	Cliff(coastal)	Deposit	Pediment	
	Catchment	Currents	Process	Pipes	
	Channel	Cusp (coastal)	Landforms	Rills	
	Cut-off	Delta (coastal)		Scree	
	Floodplain	Dune		Deposit	
	Gorge (fluvial)	Estuary		Process	
	Gully	Headland		Landforms	
	Levee	Lagoon			
	Meander	Nearshore			
	Palaeochannel	Notch			
	River	Offshore			
	River terraces	Platform			
	Valley	Saltmarsh			
	Waterfall	Spit			
	Deposit	Tombolo			
	Process	Deposit			
Landforms	Process				
	Landforms				

5. INSPIRE (blue) contains the tables ‘NaturalGeomorphologicFeatureType’ and ‘AnthropogenicGeomorphologicFeatureType’ which hold a classification system based on the draft INSPIRE data specification on geology D2.8.II.4_v2.0.1. [4]. They are linked to the ‘Level2’ and ‘Level3’ tables allowing a relationship between our classification and the INSPIRE classification.

6. ‘Geographical features’ (yellow) contains several tables that incorporate additional boundary data such as natural reserves, National Trust properties and Sites of Special Scientific Interest (SSSI). These geometries allow additional spatial queries when using the web mapping application.

After the design phase, a PostgreSQL server with PostGIS extension database software was chosen to implement the database schema. This was chosen after comparing MySQL and PostgreSQL for interoperability with GIS desktop software and web mapping libraries, spatial functions, general functions and usability [1] [2] [11].

3. Database front-end

Application development included an administrative “front-end” to ease database management, with a particular focus upon bibliographic reference creation, editing and deletion (Figure 2). The client application is built using the ExtJs4.1 Javascript framework [3] and the OpenLayers 2.11 API [8], both available under an open source licence. The server code to communicate with the PostgreSQL database uses the PHP language. The Smarty PHP Template engine is used to generate the GeoJson file for map visualization [7].

Bibliographic records | INSPIRE

ID	AUTHOR	YEAR	TITLE	PUBLICATION	PUBLISHER
1029	Al Farraj, A.,Har...	2010	Influence of hills...	Zeitschrift für G...	
1030	Chiverrell, R. C...	2008	Late Holocene e...	Geomorphology	
1031	Cundiff, P. R.	1976	Late Flandrian v...	Transactions of ...	
1032	Harvey, A.M.	1991	The influence of ...	Earth Surface Pr...	
1033	Harvey, A.M.	1992	Process interact...	Geomorphology	
1034	Harvey, A. M.	1997	Coupling betwe...	Geomorphologie...	
1035	Harvey, A. M.	2001	Coupling betwe...	Catena	
1036	Harvey, A. M.	2002	Effective timesc...	Geomorphology	
1037	Harvey, A. M.	2007	Differential reco...	Geomorphology	
1038	Harvey, A. M.	2007	Geomorphic inst...	Geomorphology	
1039	Harvey, A. M.;O...	1981	Dating of post-q...	Earth Surface Pr...	

Page 1 of 8

Abstract for ART_ID=1036

This paper presents a review of the coupling concept in fluvial geomorphology, based mainly on previously published work. Coupling mechanisms link the components of the fluvial system, controlling sediment transport down the system and the propagation of the effects of base-level change up the system. They can be viewed at several scales: at the local scale involving within-hillslope coupling, hillslope-to-channel coupling, and within-channels, tributary junction and reach-to-reach coupling. At larger scales, coupling can be considered as zonal coupling, between major zones of the system or as regional coupling, relating to complete drainage basins. These trends are illustrated

Update

Levels Images Spatial

Delete Refresh connection

Level1	Level2	Level3
Timescale or p...	Century	dummy
Environment/ ...	Fells	dummy
Environment/ ...	Slope	Gullies
Environment/ ...	Fluvial	Channel
Attributes with	Crunlin	Alumny

Level selection

Level1: Environment/ Landform

Level2: Glacial

Level3: Moraine

Add

Record details

ID: 1036

Authors: Harvey, A. M.

Year: 2002

Title: Effective timescales of coupling in fluvial systems

Publication: Geomorphology

Publisher:

Start page: 175

End page: 201

Volume: 44

Year 2:

Link:

DOI:

Place: Howgill Fells

NCA: Howgill Fells

Source: journal

map Places Update record

FIGURE 2

The front-end allows the insertion of all the data necessary to define a bibliographic reference: it is possible to upload images and maps, insert an abstract and create the reference bounding boxes using an interactive map

(Figure 3).

In addition, the bulk-upload of references from Endnote (using the XML file format) is supported; therefore the data manager can collect a large group of references and abstracts and upload them as a batch job to the database.

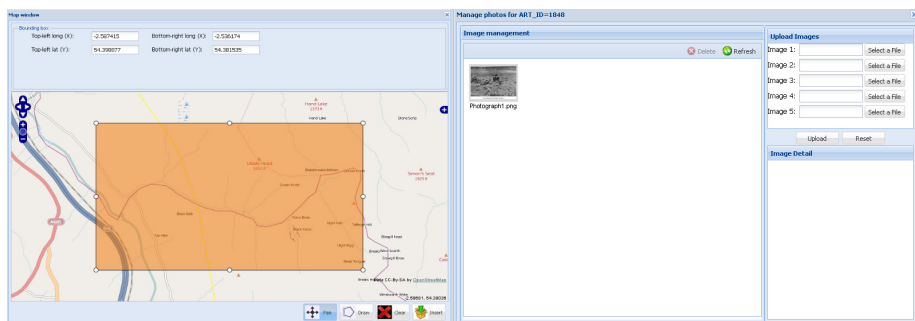


FIGURE 3

4. Web mapping application

A fully interactive web mapping application allows users to perform both spatial and aspatial queries (Figure 4).

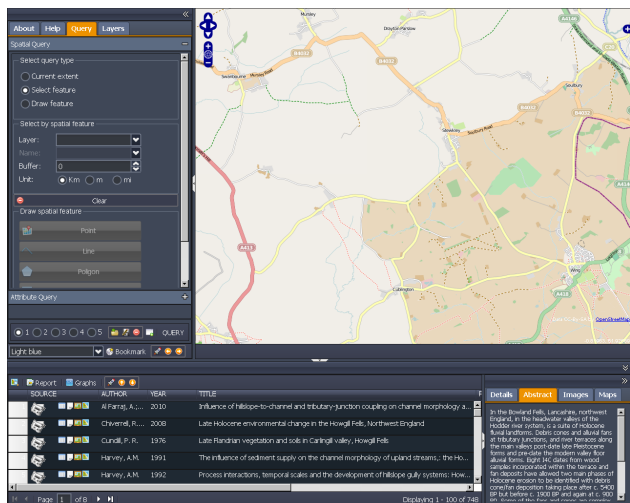
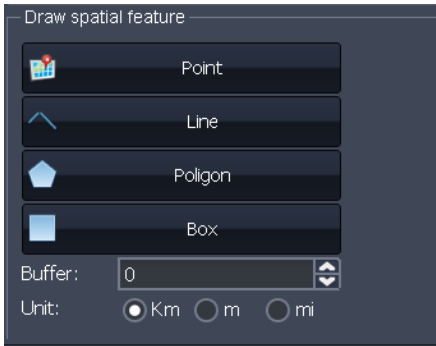


FIGURE 4

FIGURE 5



The user can carry out spatial queries using the current map extent, choose the boundary of features stored inside the database or draw a custom feature on the map (Figure 5).

Additional aspatial query parameters include the level 1-2-3 classification, reference attributes and an abstract keyword (Figure 6).

The application allows saving up to 5 queries using the html5 localstorage API, for older browsers the query can be saved in text format (Figure 7).

FIGURE 6

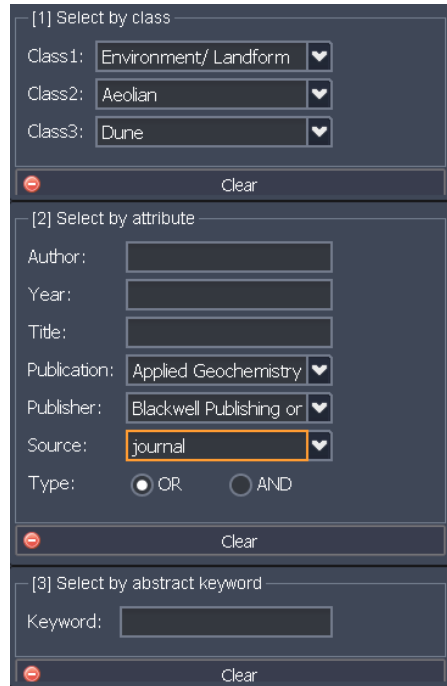
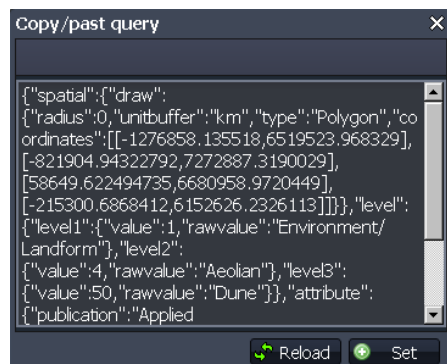
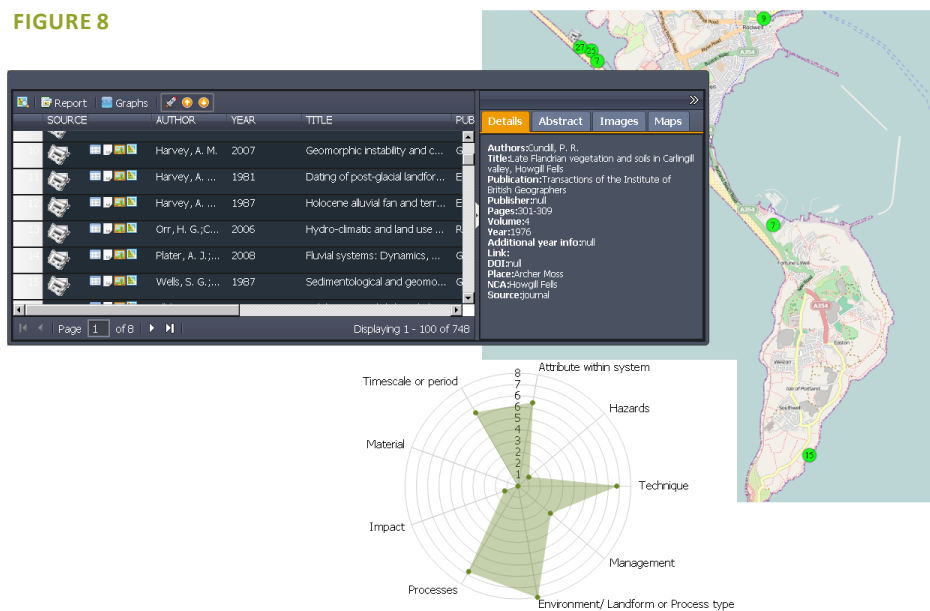


FIGURE 7



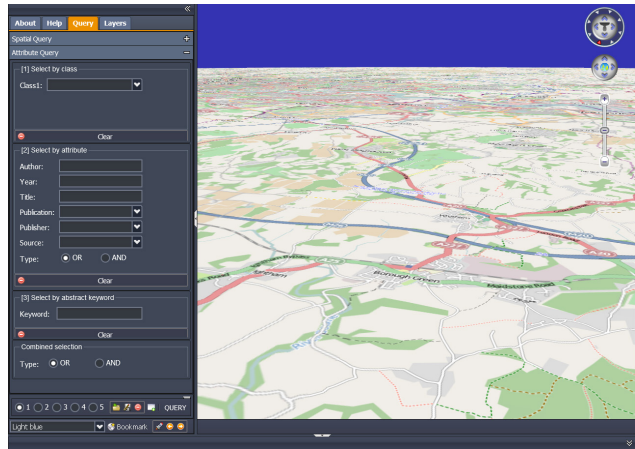
Query results are visualized both on a map and on a grid (x§). Additional information such as graphs and images are visualized for each selected reference, moreover the user can download a custom report containing a map and the reference details.

FIGURE 8



At small scales, features overlap and a map becomes cluttered; features are therefore displayed as clusters. Clustering is possible on the client browser but only for a small number of features. Clustering hundreds of features is not a feasible option because the performance of the client user interface rapidly degrades, for this reason at small scales data is clustered using SQL queries before being sent to the client.

Future development includes a mobile version built with the open source Touch2 framework and a 3D version built with the open source OpenWebGlobeSDK [9] (Figure 9). 3D mapping will be useful to allow a more realistic visualization of landscapes and landforms and the associated references.

FIGURE 9

5. Conclusions

This project has focused upon the development of a spatial database and accompanying database administration and web mapping applications for the provision of reference information on the geomorphology of the British landscape to professional end-users and the general public. Initial focus has been on the compilation of reference information for two areas, including part of the High Speed 2 (HS2) rail link between London and the West Midlands. Reference sources include journal articles, government reports, books and PhD dissertations.

The client front-end allows easy administration of records in the underlying database, including the bulk import of references from Endnote XML files. Additionally, a bespoke mapping tool allows the insertion and modification of bounding boxes together with other table attributes. A three-level classification system allows category “tagging”, aiding end-users in database searches. The web mapping application is currently under development using open source software and it will be completed during summer 2012.

On completion, the project will have completed technical development and compiled reference information, in detail, for two sample areas. Future work will look to compile full reference information for the United Kingdom and Northern Ireland, further helping to identify gaps in knowledge and help direct future geomorphological research.

[1] Compare SQL Server 2008 R2, Oracle 11G R2, PostgreSQL/PostGIS 1.5 Spatial Features. Boston GIS website. Retrieved May 7, 2012 from http://bostongis.com/?content_name=sqlserver2008r2_oracle11gr2_postgis15_compare#221.

[2] Cross Compare SQL Server 2008 Spatial, PostgreSQL/PostGIS 1.3-1.4, MySQL 5-6. Boston GIS website. Retrieved May 7, 2012 from http://bostongis.com/?content_name=sqlserver2008_postgis_mysql_compare#178.

[3] ExtJS/Touch2 frameworks. Sencha website. Retrieved May 7, 2012 from <http://www.sencha.com/>.

[4] INSPIRE, D2.8.II.4 Data Specification on Geology - Draft Guidelines, 2011. INSIRE website. Retrieved May 7, 2012 from http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GE_v2.01.pdf.

[5] MySQL vsPostgreSQL. WikiVS website. Retrieved May 7, 2012 from http://www.wikivs.com/wiki/MySQL_vs_PostgreSQL.

[6] National Character Areas. Natural England website. Retrieved May 7, 2012 from <http://www.naturalengland.org.uk/ourwork/landscape/englands/character/areas/default.aspx>.

[7] OBE, R., AND HSU, L. *PostGIS In Action*. Manning, Stamford, 2011.

[8] OpenLayers API. OpenLayers website. Retrieved May 7, 2012 from <http://openlayers.org/>.

[9] OpenWebGlobe SDK. OpenWebGlobe website. Retrieved May 7, 2012 from <http://wiki.openwebglobe.org>.

[10] OS OpenData. Ordnance Survey website. Retrieved May 7, 2012 from <http://www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html>.

[11] PostgreSQL vs. MySQL, Pros & Cons. IT Management & Leadership eMagazine website. Retrieved May 7, 2012 from <http://leadit.us/hands-on-tech/PostgreSQL-vs-MySQL-Pros-Cons>.

[12] How to Graph data in PostgreSQL. Postgres Online Journal website. Retrieved May 7, 2012 from http://www.postgresonline.com/journal/archives/188-plr_part1.html.

[13] WESTRA, E. *Python Geospatial Development*. Packt Publishing, Birmingham, 2010.

Using GRASS and PostgreSQL/PostGIS for the development of the automatic preprocessing for a distributed vector-based hydrological model

Authors

- **Sonja Jankowsky** , IRSTEA, UR HHLY, 69336 Lyon, LUNAM Université, IFSTTAR, GER, 44341, Bouguenais, *France*
- **Pedro Sanzana**, University of Chile, FCFM, Av. Blanco Encalada 2120, Santiago, *Chile*
- **Flora Branger**, IRSTEA, UR HHLY, 69336 Lyon, *France*
- **Isabelle Braud**, IRSTEA, UR HHLY, 69336 Lyon, *France*
- **Yvan Paillé**, IRSTEA, UR HHLY, 69336 Lyon, *France*
- **Florent Brossard**, IRSTEA, UR HHLY, 69336 Lyon, *France*
- **Fabrice Rodriguez**, LUNAM Université, IFSTTAR, GER, 44341, Bouguenais, *France*

KEYWORDS : Distributed hydrological modeling, vector-based model mesh, spatial discretization, preprocessing

Introduction

Many GIS tools have been developed for the preparation of the geographical input data of distributed hydrological models such as the hydrological sciences toolbox of GRASS [1], the raster terrain analysis plug-in of QGIS [2], the Taudem [3] extension of Mapwindow, the terrain analysis hydrology package of SAGA [4], etc. However, most of these preprocessing tools consider raster data, not vector data. Model meshes based on irregular vector geometries are mainly used by [5], [6] and [7]. Lagacherie et al. ([5]) developed the landscape discretization tool GEOMHYDAS in GRASS GIS, which

allows the construction of vector-based HRU maps (Hydrological Response Units) based on a selective overlay of property layers such as land use, soil maps or sub-basins. Furthermore, special tools were developed to integrate man-made features such as ditches, roads and agricultural fields. Branger [6] simply used a land use map as model mesh. Rodriguez et al. [7] developed the preprocessing of the urban hydrological model URBS in MapInfo. Instead of HRUs, URBS uses Urban Hydrological Elements (UHEs) as model units, each consisting of one cadastral unit and half of the adjoining street.

In this paper, we go one step further by combining the HRU model mesh with UHEs in order to create a model mesh adapted for peri-urban areas. This particular model mesh needed the development of several scripts for the extraction of the flow routing in a mixed urban and rural environment. In order to obtain realistic flow paths, mesh optimization methods were developed treating concave or too large polygons or polygons with holes. Furthermore, scripts were developed allowing the integration of raster information (e.g. slope) into the model mesh without creating a large number of small polygons.

Software requirements

Following the experience of Branger et al. [8], who tested different open-source GISs such as OpenJump, SAGA, OrbisGIS, Mapwindow, PostGIS, QGIS and GRASS for the preprocessing of distributed hydrological models built within the LIQUID modeling framework, we chose GRASS GIS (<http://grass.fbk.eu/>) because of its topological data structure and functions for the development of the automatic preprocessing. However, some tasks, especially concerning the flow routing, have been shown to be easy to develop using PostGIS functions embedded in SQL queries and more complicated to develop using GRASS functions. Therefore, GRASS functions and PostgreSQL queries were integrated into Python scripts using the Pygresql and Grass libraries.

Developed method

UHE and HRU delineation

In a first step, the preprocessing developed by Rodriguez et al. [7] in MapInfo

for urban zones (UHEs) was transformed into GRASS-Python scripts. These scripts [9] allow additionally to create the real UHE geometry by dividing the road in halves and merging it with the adjoining cadastral unit.

In rural areas, the HRUs can be created with the `m.seg/m.dispolyg` GRASS script [5] using different polygon property layers and the drainage network as input data. The UHEs and HRUs are then overlaid using GRASS `v.overlay`.

In order to integrate raster properties, the slope segmentation GRASS-Python script was developed [10]. The script takes as input the HRU vector map, a slope raster map and a threshold value and extracts all polygons for which the standard deviation of the slope is larger than the threshold value. The extracted polygons are divided along two morphological boundaries defined by the inter quartile range. This range, defined as the difference between the third quartile (Q_3) and the first quartile (Q_1), allows the identification of homogeneous classes (Figure 1). The intersection lines are smoothed using the Snakes and Douglas-Peucker algorithm implemented in GRASS GIS, see Figure 1. The script works also with any raster property in addition to slope, such as elevation, aspect, etc.

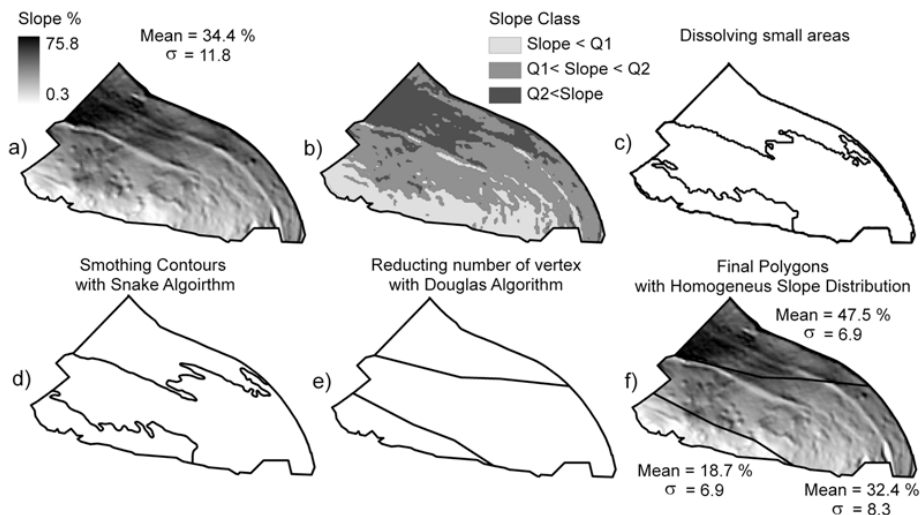


FIGURE 1

Segmentation of a polygon using homogenous slope classes [10]

Three scripts were developed in order to adapt the model mesh to numerical constraints [10] :

1. The polygon_hole GRASS-Python script divides polygons with islands along the nearest and farthest distance from the island to the surrounding polygon.
2. The convexity segmentation script extracts concave polygons for which the convexity index (perimeter of a convex polygon with the same area/concave polygon) exceeds a certain threshold value and partitions them using a C++ triangle algorithm [11]. The triangle algorithm was preferred to the v.delaunay triangulation available in GRASS as it creates larger and less long-drawn polygons. In order to reduce the number of model units and to keep the object oriented character of the irregular model mesh, the triangles were re-unified with a GRASS-Python script under the constraint of a convexity threshold.
3. The area segmentation script is similar to the convexity segmentation script, but uses a maximal polygon area as threshold value.

Flow transfer and routing

For the peri-urban areas typical flow transfer, three principles were applied:

1. The overland flow paths follow the topography towards the lowest neighbor model unit.
2. The surface runoff on impervious areas is intercepted by sewer pipes. Therefore, the flow transfer from the UHEs to the closest drainage network is direct.
3. The lateral subsurface flow depends on the water table level and can be multi-directional towards all neighbors.

Three PostgreSQL/GRASS-Python scripts were developed for the determination of these flow transfer paths [12]. Scripts 1 and 2 are based on recursive functions, which extract single flow directions between model units. The overland flow paths are determined sub-basin by sub-basin, whereas the urban surface runoff connections are independent of the sub-basins. The polygon-polygon neighborhood relations in script 3 were extracted with GRASS functions. However, PostgreSQL queries had to be used to extract

the neighborhood relations of model units (polygons) and river reaches (lines). For this, the polygons were transformed into boundaries. A small buffer around the centroid of the boundaries was then intersected with the river reaches in order to determine only the adjacent river reach.

Concerning the drainage network three scripts were developed :

1. The script `river_direction` orientates all network reaches in upstream direction.
2. The script `num_river` numbers all network reaches starting from the outlet.
3. The script `river_h_s` calculates the mean altitude and slope per network reach.

Results

Thanks to the developed scripts, the model meshes and flow routing for two small catchments (Chaudanne and Mercier, respectively 4.1 and 6.8 km²) in the peri-urban area of Lyon city, France, could be determined as shown in Figure 2 for the Chaudanne catchment. On this catchment, we obtained 2945 HRUs with average area of 1394 m² and 522 drainage reaches.

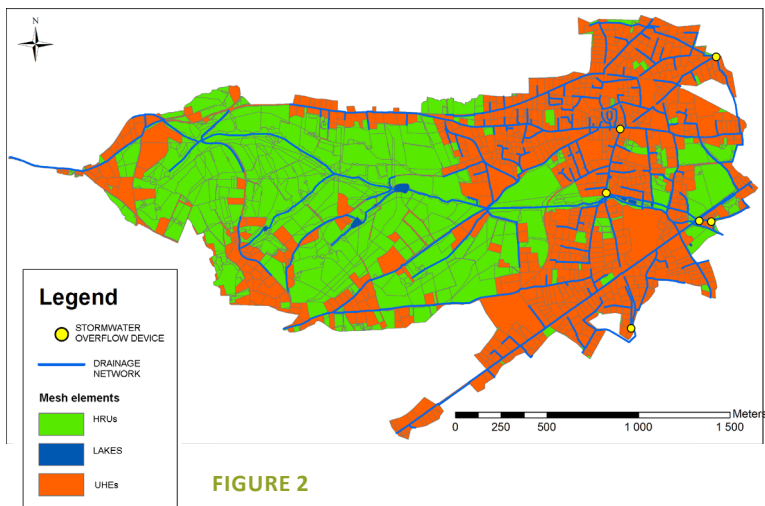


FIGURE 2

Final model for the Chaudanne catchment consisting of UHEs (red) and HRUs (green)

The mesh optimization improved considerably the flow paths and the area difference between the model units, as can be seen in Figure 3. Before the mesh optimization (Figure 3a) some of the flow paths cross the river network due to large and concave polygons, which could be corrected by means of the mesh optimization (Figure 3b).

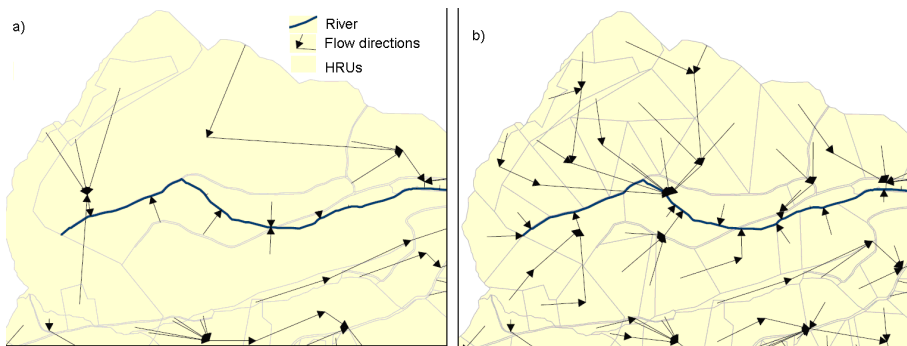


FIGURE 3

Determined overland flow paths for a) the non-optimized mesh and for b) the optimized mesh

Encountered problems

The study revealed that not all tasks can be completed automatically. Especially polygon-line relations are not well represented in current GIS software and they often need manual treatments. Furthermore, no open-source GIS function could be found determining the middle street line. Therefore, we corrected manually the street line vectors from BD© Topo (IGN). A proper vector topology is necessary for most of the steps. We solved this problem by applying the cleaning functions of GRASS (v.clean) to the input layers and to the output layers of the major preprocessing steps.

Conclusions and perspectives

This new model mesh combining UHEs and HRUs as irregular vector geometries is promising for object oriented modeling as the catchment area

is composed of many single landscape objects. Jankowfsky [13] used these model meshes to apply the object-oriented Peri-Urban Model for Landscape Management (PUMMA). The scripts are open-source and can also be applied for other hydrological models.

Nevertheless, the irregularity of the model mesh as well as the different kind of drainage in peri-urban catchments (overland flow, pipe drainage, subsurface flow) reveal many questions concerning the runoff routing. Some of these problems, such as the correction of concave or too large polygons, as well as the integration of slope information were treated in this study. Other problems involving long polygons (e.g. roads or hedgerows) acting as flow barriers during the calculation of the flow routing still need to be addressed. A possibility would be to use the convexity segmentation script and replace the convexity index as threshold with a compactness index. Furthermore, segmentation with altitude classes similar to the slope segmentation would improve the flow routing, which is governed by the altitude differences of adjacent model units.

The developed scripts should be available soon for download at [14] and can be obtained from the authors in the meantime. In order to run the scripts, PostgreSQL 8.3/PostGIS (pginstaller [15]), GRASS [1] and PyGreSQL 4.0 [16] have to be installed.

[1] GRASS development team, 2012. GRASS Wiki: Hydrological sciences. Retrieved August 30, 2012 from http://grass.osgeo.org/wiki/Hydrological_Sciences

[2] Quantum GIS development team, 2012, Quantum GIS. Retrieved August 30, 2012 from <http://www.qgis.org>

[3] David Tarboton Hydrology Research group Utah State University, 2012, Terrain Analysis Using Digital Elevation Models (TAUDEM). Retrieved August 30, 2012 from <http://hydrology.usu.edu/taudem/taudem5.0/index.html>

- [4] SAGA development group, 2012. SAGA-GIS, Terrain Analysis - Hydrology. Retrieved August 30, 2012 from http://sourceforge.net/apps/trac/saga-gis/wiki/ta_hydrology
- [5] LAGACHERIE, P., RABOTIN, M., COLIN, F., MOUSSA, R., and VOLTZ, M. Geo-mhydas: A landscape discretization tool for distributed hydrological modeling of cultivated areas. *Computers & Geosciences*, 36(8) (2010), 1021 – 1032.
- [6] BRANGER, F. Utilisation d'une plate-forme de modélisation environnementale pour représenter le rôle d'aménagements hydro-agricoles sur les flux d'eau et de pesticides. Application au bassin versant de la Fontaine du Theil (Ille-et-Vilaine). PhD thesis, INP- Grenoble.2007.
- [7] Rodriguez, F., Andrieu, H., and Creutin, J. Surface runoff in urban catchments: morphological identification of unit hydrographs from urban databanks. *Journal of Hydrology*, 283 (2003), 146–168.
- [8] BRANGER, F., JANKOWFSKY, S., VANNIER, O., VIALLET, P., DEBIONNE, S., BRAUD, I. Use of open-source GIS and data base software for the pre-processing of distributed hydrological models, *Geospatial free and open source software in the 21st century*, Bocher, E., Neteler, M. (Eds), Lecture notes in Geoinformation and Cartography, 2011, 12 pp.
- [9] PAILLÉ, Y. Automatisation du prétraitement des données spatiales pour la modélisation hydrologique distribuée en zone périurbaine à l'aide du logiciel libre GRASS. 2010. Master thesis, Université de Nantes.
- [10] Sanzana, P., JANKOWFSKY, S., BRANGER, F., BRAUD, I., VARGAS, X., HITSCHFELD, N., GIRONÁS, J. Automatic mesh optimization based on Hydrological Response Units for object-oriented distributed hydrological modeling, 2012, submitted to *Computers and Geosciences*.
- [11] SHEWCHUCK, J. Applied Computational Geometry: Towards Geometric Engineering, volume 1148, chapter Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator, 1996, pages 203–222.
- [12] BROSSARD, F. Automatisation du pré-traitement des données spatiales pour la modélisation hydrologique distribuée en zone périurbaine. 2011. Rapport de stage, EPMI Cergy.
- [13] JANKOWFSKY, S. Understanding and modelling of hydrological processes in small peri-urban catchments using an object-oriented and modular distributed approach. Application to the Chaudanne and Mercier sub-catchments (Yzeron catchment, France). PhD thesis, Université de Grenoble, 2011.

An open and powerful GIS data discovery engine

Authors

- **Martin Ouellet**, Library, Université Laval Québec, QC Canada, *Canada*
- **Stéfano Biondo**, Library, Université Laval Québec, QC Canada, *Canada*

KEYWORDS : semantic metadata mining, thesaurus, combined text-spatial search

The Geographic and Statistical Information Center (Centre GéoStat) of Laval University's Library is responsible for the acquisition, the organization and the dissemination of geospatial data. The center's objective is to support all students, teachers, and researchers across campus in their teaching and research work. The center manages more than 30 000 datasets, which represent about 10To of hard drive space. Like most academic libraries, its collection consists of raster and vector dataset useful for many different fields such as geography, political sciences, forestry, geology, health, soil science, marketing, history, topography, etc.

From a technological point of view, the "Big Data" phenomenon, of which geospatial data is part, will surely prove to be one of the landmarks of the 2010-2020 decade [2]. While data keeps growing, hardware gets cheaper and cheaper. Therefore, we have the means to store in-house entire collections of data. However, most organizations postpone the implementation of a discovery engine and leave unexploited this wealth of information.

In the past, geospatial data was associated only to a few specific disciplines, such as geography, engineering or land surveying. However, during the last decade, the use of geospatial data have spread because it became widely accessible for free (Open Data) and because more tools were developed to create, gather, exploit and disseminate this type of data (Google Earth, Open Source

software, « Mashup », iPhone and other GPS-enabled mobile technologies) [7]. The number of users grew exponentially and got more and more diverse (this can also be observed in academic settings). The statistics compiled by the Geographic and Statistical Information Centre show that today's clientele transcends disciplines. The new groups of users are generally inexperienced and are in need geospatial data for various and "untraditional" reasons: to back sociological or demographic studies, to plan itineraries, to help managers or company administrators take decisions, etc. In a nutshell, clientele is on the rise, clearly divided between novices and experts, and geospatial data can now be used for many different purposes.

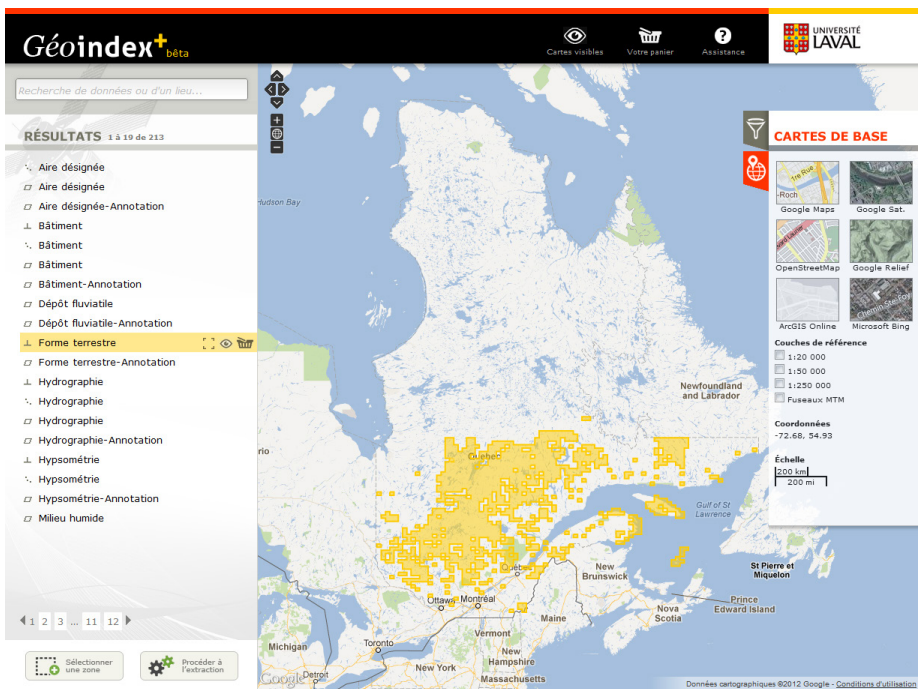


FIGURE 1

Screen capture of Geoindex+ interface

The challenge is to adapt to both realities: the ever-increasing amount of data and the changing nature of users. It is paramount to develop tools that are user friendly (to accommodate the various degrees of expertise) and powerful (to cope with the volume of data) to help users find what they need in this mass of information. Unfortunately, user-friendly search interfaces and better technologies don't always adequate with optimal outcome. Enabling semantic networks can significantly increase the relevance of search results, in accordance with users' needs. Therefore, the Centre GéoStat decided to develop a powerful discovery engine that combines traditional text-based search and spatial filter: Géoindex+ (Figure 1).

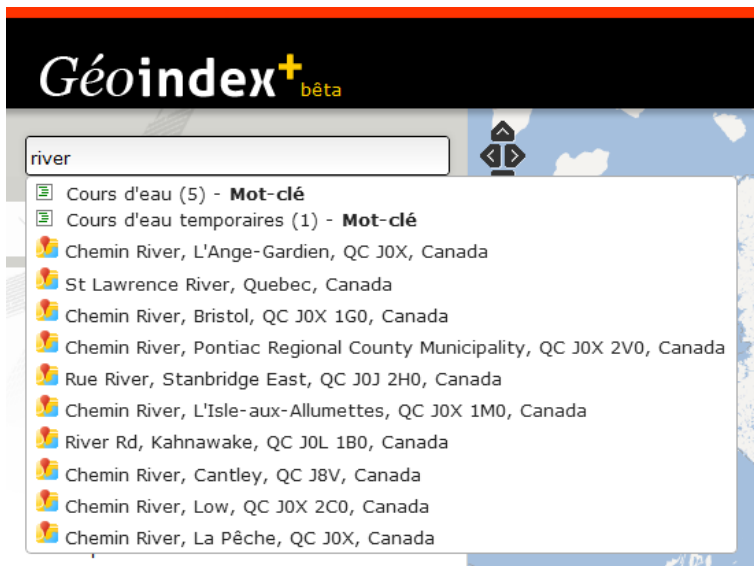


FIGURE 2

Screen capture of Geoindex+ simple search box

It features a simple search box (Figure 2) combining both metadata queries (to find relevant dataset) and the geocoding service API from Google (to help user locate on the map). If the user chooses to navigate directly on the map, the list of result will be automatically updated based on the current map extent.

The effectiveness of any search engine rests on the richness and quality of metadata. In March 2006, the Geographic and Statistical Information Centre set up a committee consisting of geospatial data users from Université Laval. The role of this committee was to enable and foster campus-wide access to geospatial data, to facilitate expertise sharing and to make the acquisition of most data possible. Work carried out by the committee led to the creation of a metadata framework that complies with the North American Profile of ISO 19115:2003 [1]. This allows for a standardized description of the various data sets, supports more effective data discovery and improves interoperability between different search engines.

We decided to use GeoNetwork, an geographic metadata catalog application that appears to be the most advanced open source product on the market [3]. It is an application with continuous development and a growing community. It is currently used in numerous Spatial Data Infrastructure initiatives across the world such as FAO, INSPIRE, WHO, etc. [5]. GeoNetwork has given us a web-based environment for editing and storing metadata within a database.

Special care was taken during metadata entry to ensure consistency in the choice of terminology. ISO 19115 allows for the selection of any recognized thesaurus [6]. We chose an encyclopedic thesaurus instead of a more specific one because our data set is multidisciplinary and not limited to a single academic subject matter. The Répertoire de vedettes-matière (RVM), a national standard for French-language indexing since 1974, met all of our needs [4]. Developed and maintained by Laval University, this encyclopedic thesaurus contains more than 273 000 entries, enhanced by semantic networks (specific terms, related terms, cross-references and English-language equivalents). It allows us to describe metadata using the appropriate vocabulary. In addition, the RVM is flexible and continually evolving by allowing us the possibility to modify or add new terms.

Inspired by the new generation of library discovery tools, we wanted to implement faceted searching in our engine to allow users refine results according to fields specific to GIS files. We generated indexes from our

GeoNetwork metadata records through the Apache Solr platform [9]. This powerful indexing technology has proven with large volumes and can easily support queries over millions of records. Considering that we will have a few thousand products listed in our system, the performance of the text search through our metadata is more than sufficient. This technology also made it possible to add to our discovery engine a “semantic autocomplete” search box and to enable full-text searching.

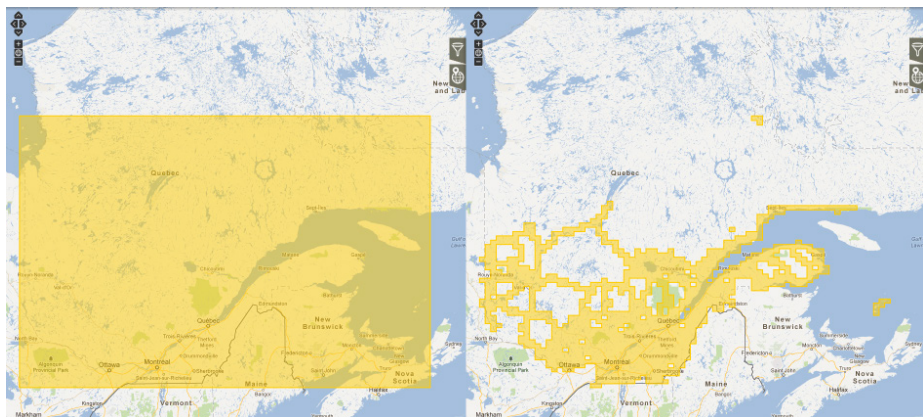


FIGURE 3

Default minimum rectangle versus more accurate complex polygon

ISO 19115 includes a section pertaining to the geographical scope of data sets. This scope is defined by the minimum bounding rectangle capturing the whole of the data set. Although essential to the formulation of a basic spatial search, the delineation of these boundaries can be considered too vague. In order to alleviate this problem, we created more precise and more complex delimitations for each data set, this to represent more faithfully the territory covered (Figure 3). These complex polygons were created using geometric operators from the PostGIS database [8]. Then we associated every metadata record to its geometric shape. It is therefore possible to refine a search by

moving across the map. We would like to keep this complex geometry directly in the metadata description instead of the default minimum rectangle because it is more relevant for the spatial filter. The XML implementation of the ISO19115 standard provides a definition for this type of complex polygon (in GML format), but it is not fully supported in the version of GeoNetwork that we use (2.6.3). To ensure that our solution is effective (at the end with the entire inventory that we have), we have simulated queries (with both spatial and textual filter) with nearly 3,000 metadata records combined with the same number of complex polygons. The response time was never more than a few seconds in every case.

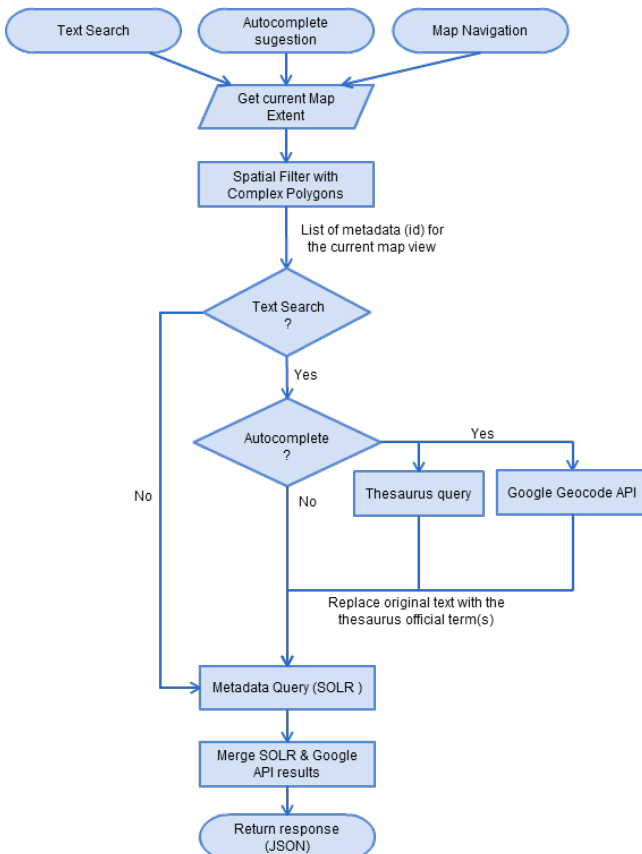
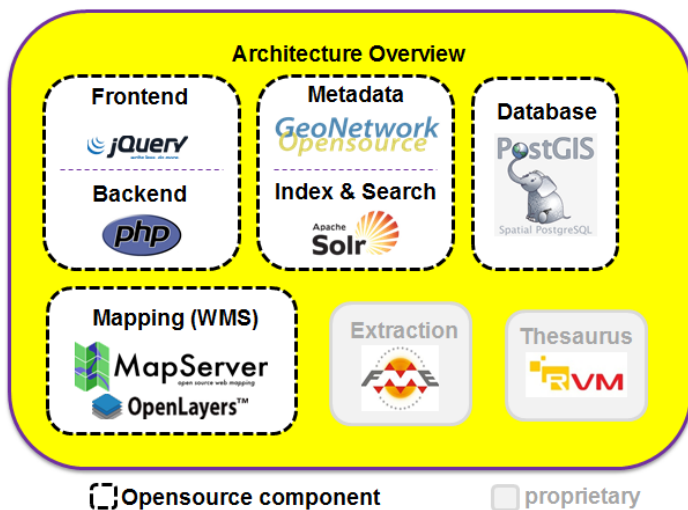


FIGURE 4
The search workflow

The search engine (Figure 4) we developed meets the challenge of helping various users to find what they need in a sea of data. It is also the corner stone of a bigger project: a web-based platform called Géoindex+ that is used to disseminate and extract geospatial data. The search capabilities and the geospatial data visualization section of this application is an assembly of existing opensource components (Figure 5). To put forward the characteristics of the search engine, demonstrations with Geoindex+ will be performed during the communication.

FIGURE 5

Architecture Overview



This communication is aimed to everyone who has to bridge the gap between a huge amount of geospatial information and ease of discovery. Those who are convinced of the importance of metadata to facilitate discovery, but that need to find time (and courage) to address this issue, will certainly benefit from this communication.

- [1] CANADIAN GENERAL STANDARDS BOARD. North American Profile of ISO19115:2003 - Geographic information – Metadata. USA-Canada, 2008. Retrieved November 15, 2010 from http://www.cits.rncan.gc.ca/html/brodeurj/.protege/.napMetadata/review3/napMetadataProfileV1_2_1_en_final20081215.pdf
- [2] DUMBILL, E. Planning for Big Data. O'Reilly Media, Inc.,2012.
- [3] FEDERAL GEOGRAPHIC DATA COMMITTEE. ISO Metadata Editor Review, 2009. Retrieved December 2, 2009 from <http://www.fgdc.gov/metadata/iso-metadata-editor-review>
- [4] GASCON, P. Le Répertoire de vedettes-matière de la Bibliothèque de l'Université Laval : sa genèse et son évolution. *Documentation et bibliothèques* 39, 3 (1993), 129-139.
- [5] GEONETWORK OPENSOURCE. Gallery. Retrieved June 18, 2012, <http://geonetwork-opensource.org/gallery/gallery.html>
- [6] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 19115 - Geographic information, metadata, Geneva, 2003.
- [7] JONSHON, L., LEVINE, A. and SMITH, R. 2009. *The 2009 Horizon Report*. Retrieved December 18, 2009 from <http://net.educause.edu/ir/library/pdf/CSD5612.pdf>
- [8] OBE, O. R. and HSU, LEO S. *PostGIS in action*. Manning, Stamford, 2011.
- [9] SMILEY, D. and PUGH, E. *Apache Solr 3 Enterprise Search Server*. Packt Publishing, Birmingham, 2011.

Ontology Based Domain Specific Search of Crowdsourced OpenStreetMap Dataset and Wiki

Authors

- **Stefan Keller**, University of Applied Sciences Rapperswil (HSR),
Switzerland
- **Michel Ott**, University of Applied Sciences Rapperswil (HSR),
Switzerland

KEYWORDS : information retrieval, ontology, text processing, open data, crowdsourcing

Finding relevant tags in the OpenStreetMap (OSM) database is a difficult endeavor even for experienced users of this still growing, well-known crowdsourced mapping project.

Within the OSM project there exists no authoritative and often no definition whatsoever of what a tag means. Tags are equivalent to attributes or key-value-pairs and serve also as map keys in interactive maps. OSM follows a decentralized yet collaborative manner with no central authorship of OSM. The advantage of this approach is the widespread participation in authoring content [2]. But the heterogeneity and the lack of structure makes it hard to satisfy the need to get a relevant answer while seeking for a tag or tag-scheme based on few simple real-world nouns expressed in local language.

With the project 'TagFinder' we are trying to satisfy the information-hunting task to find relevant tags based on a simple text query. This is the first and only web application which tries to deliver such functionality. For example, given a search term "Restaurant" a result page is being presented

with “amenity=restaurant” as the top ranked entry. If you search the german words “Autobahn” (highway) or “Gemeindehaus” (parish hall) in you won’t find the corresponding tags in the top 10 list of the result page of OSM Wiki. And in Taginfo service of OSM [4] there are no results found whatsoever.

In order to achieve this goal, mainly the OSM Wiki is being processed as input source as well statistics about the OSM database.

Within this contribution we describe the approach we have chosen in order to resolve this domain specific information retrieval task (Figure 1).

Result

Search String Analysis	
Search String:	Restaurant
Detected Language:	de
Preferred Value:	restaurant

FIGURE 1

Information flow control of a user query within Tagfinder

Count	Key	Value	Filter
247604	amenity	restaurant	[+]
184	building	restaurant	[+]
158	bak:fac_type2	RESTAURANT	[+]
134	wpt_symbol	Restaurant	[+]
124	kvl_hro:amenity	restaurant	[+]
100	designation	Restaurant	[+]
90	sym	Restaurant	[+]
69	amenity	parking;restaurant;fuel	[+]
69	name:en	Bada Fish Restaurant	[+]

TagFinder has been implemented within a web application which offers industry-standard compliant webservice for Point-of-Interests from OSM [1] and is available for testing purposes (see [4]).

Given the running web application the user enters a query. First, the language of the input search term is determined by a thesaurus as a part of a domain-specific ontology and translated into an English equivalent. Thus, it is basically not important, in which language the end user enters the query. The thesaurus, which was developed especially for the OSM nomenclature, is using the Simple Knowledge Organization System (SKOS) schema [3]. SKOS is very suitable for the development of word lists, links, and their translations. In SKOS, there are two basic types of terms. On the one hand, the «concept scheme», which serves as a high-level group element, and the «Collection» for the collection of semantically related terms. The «concept scheme» is a defined controlled vocabulary such as the «Eating place», which subsumes all the possible terms of Restaurant. On the other hand there is the «concept», which is the smallest element in a SKOS thesaurus. A «Concept» describes a single term, including all its dependencies and translations. Thus, for example, concepts have father-son relationship or a reference to a preferred term.

With the help of the English equivalent, multiple requests are made on the Taginfo service [4] to find out which key-value pairs most likely corresponds to the term the user entered. Then the list will be filtered in order to eliminate duplicates. Then it is sorted in descending order of frequency and output (Figure 2).

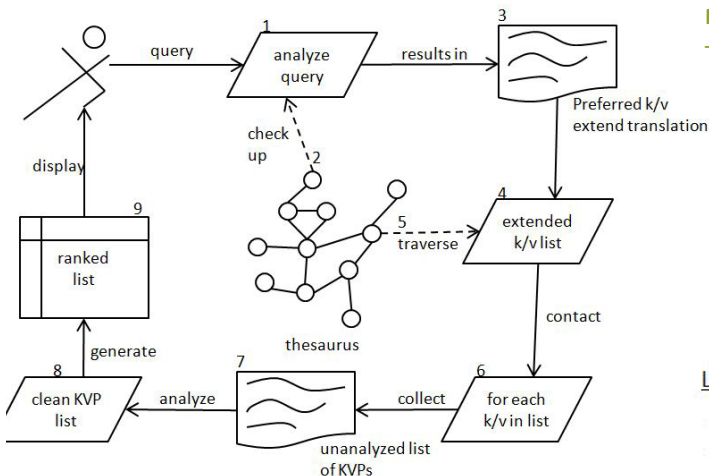


FIGURE 2
TagFinder result page

Legend:
--> reading data
-> action

TagFinder offers also an application programming interface (API) so that it can be used by other applications like JOSM, the OSM editor.

Future work of TagFinder includes an extension of the database to cover the whole world as well as more sophisticated preprocessing and matching functions. There are several enhancements in consideration, like the management of the thesaurus. Here TagFinder is based on crowdsourcing principles itself. It uses special entries (a kind of templates) from OSM Wiki in order to indicate related terms. Then, in the preprocessing phase, we hope to enhance the result quality by filtering out proper names (like 'Windsor Castle'). The ranking can be further tuned depending on the fact, if the search term occurs in the key or value part of a tag candidate, and if there exists a Wiki page for the tag candidate. Finally, if there are too few results, fuzzy linguistic string matching functions, like trigrams, can support the retrieval of relevant tags given simple search terms.

[1] AUER, M., AND ZIPF, A. OpenStreetMap meets Open Geospatial Consortium - Zum Anwendungspotential einer Geodateninfrastruktur mit freiwillig gesammelten Geodaten. In Schilcher, M. (Hrsg.): Geoinformationssysteme. Beiträge zum 16. Münchner Fortbildungsseminar, 2011, pp. 106-113.

[2] CHAKRABARTI, S. Mining the Web - Discovering Knowledge from Hypertext Data. Morgan-Kaufmann Publishers, 352 p. (2002). ISBN 1-55860-754-4.

[3] SKOS, 2004. SKOS Simple Knowledge Organization System. Retrieved July 15, 2012 from <http://www.w3.org/2004/02/skos/>

[4] TAGFINDER, 2012. OpenStreetMap TagFinder Documentation. Retrieved July 15, 2012 from <http://giswiki.hsr.ch/TagFinder>

[5] TOPF, J. 2012. Taginfo API Documentation. Retrieved July 15, 2012 from <http://taginfo.openstreetmap.org/taginfo/apidoc>

An open tool to register landscape oblique images and generate their synthetic model

Authors

- **Timothee Produit**, Ecole Polytechnique Fédérale de Lausanne, *Switzerland*
- **Devis Tuia**, Ecole Polytechnique Fédérale de Lausanne, *Switzerland*

KEYWORDS : Space resection, monoplottter, glacier survey, python

Introduction

Thousands of landscape images stored in archives or in the web are not georeferenced or have a rough georeference (inaccurate 2D location, location name). However, quantitative studies based on images require the precise pose (location and orientation) of the camera. Monoplotting softwares use a Digital Elevation Model (DEM) to assign world coordinates to each pixel of an image. The collinearity equation (Equation 1) transforms the world coordinates in image pixel coordinates. It is used to plot GIS vectors in the image plane. Its inverse function is used to geo-rectify the landscape image. The monoplottter implementation presented in this paper is based on opensource softwares.

User-defined Ground Control Points (GCP) are the common way to register an image. GCP identify corresponding features in the image and in the ortho-image. They are exploited in several ways. First, georeferencing process rectifies the landscape image and maps it on the ortho-image using local transformations between the GCP. However, georeferencing is more suited for aerial nadir images. Its application on landscape image needs a careful and

time demanding digitization of many GCP around the study area [1]. Image portions far from the GCP are highly distorted. Second, GCP are exploited as 2D – 3D correspondences, the third dimension is interpolated from the DEM. 2D-3D GCP are used to compute the precise pose of the camera [2]-[7]. Once the pose has been computed, a world coordinate is attributed to each image pixel. Monoplotter softwares have been used to study glaciers [8], mountain landscapes [9] or as teaching softwares for photogrammetry [10].

Monoplotters usually provide the following functions :

- Camera pose estimation based on 2D-3D correspondences;
- Image ortho-rectification;
- Image vector to GIS vector transformation;
- GIS vector to image vector transformation (widely used in Augmented Reality).

The main addition of our software is the computation of a synthetic image, which is the projection of the ortho-image in the image plane. Ortho-image rendering on a DEM is a common 3D-GIS task. Our software offers an open implementation, which does not make use of the graphic card. It focuses on the rendering quality rather than on rendering rapidity.

In the following section we will introduce the mono-plotting concept, the software functions and implementation. Finally, we will show some possible platform usage, based on ancient pictures of the Aletsch glacier (Southern Switzerland).

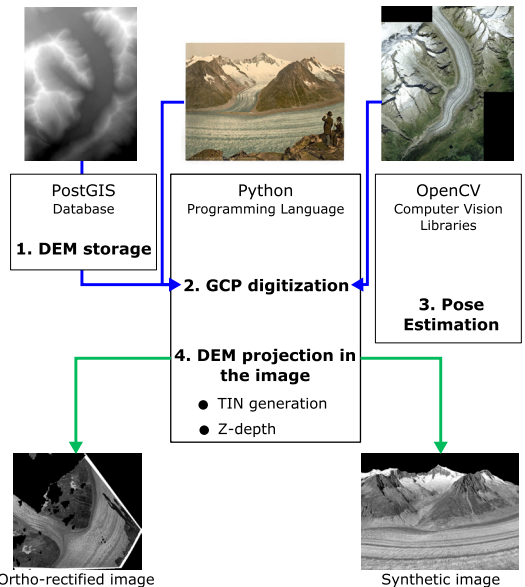


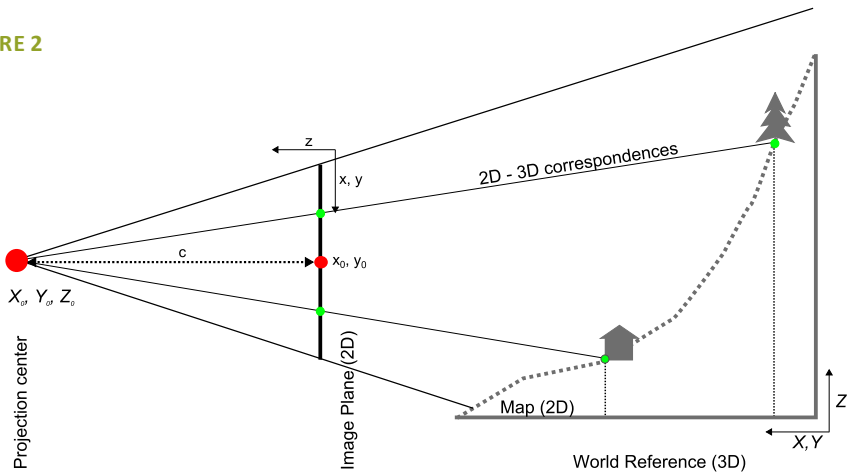
FIGURE 1

Our monoplotting software

Mono-plotting

Given a camera pose, the monoplotting software computes the transformations between the image and the 3D model. Figure 2 illustrates the mono-plotting concept. The camera pose is described by the camera location (red point) and orientation. The location (projection center) has coordinates X_0 , Y_0 and Z_0 . The camera orientation is the angle between the viewing direction along the focal c and the X , Y plane. In 3D, the camera orientation is described by 3 angles. Each pixel has a correspondence in the 3D model. The projection of the 3D coordinates on the map generates GIS layers (ortho-image and GIS vectors). The projection of an ortho-image in the image plane generates a synthetic image, while projecting GIS vectors generates semantic information.

FIGURE 2



Pose estimation

The pose estimation is the central step of the monoplotting. The 2D – 3D correspondences respectively digitized in the image and in the ortho-image are used to solve the collinearity equation (Equation 1).

In this equation c is the focal, expressed in [mm] or in [pixel]. \mathbf{x} and \mathbf{y} are the pixel coordinates. $\mathbf{x0}$ and $\mathbf{y0}$ are the coordinates of the principal point (usually the center of the projected image). These parameters describe the camera intrinsic parameters. Extrinsic parameters are $\mathbf{X0}$, $\mathbf{Y0}$ and $\mathbf{Z0}$, (the camera coordinates) and orientation of the camera described by a rotation matrix \mathbf{R} . If the camera intrinsic parameters are unknown, the focal [pixel] is computed from the focal [mm] stored in the image metadata and camera specification found on the web. The principal point is assumed to be at the image center and no image plane deformations are taken into account.

EQUATION 1

$$x - x_0 = -c \cdot \frac{r_{11} \cdot (X - X_0) + r_{21} \cdot (Y - Y_0) + r_{31} \cdot (Z - Z_0)}{r_{13} \cdot (X - X_0) + r_{23} \cdot (Y - Y_0) + r_{33} \cdot (Z - Z_0)}$$

$$y - y_0 = -c \cdot \frac{r_{12} \cdot (X - X_0) + r_{22} \cdot (Y - Y_0) + r_{32} \cdot (Z - Z_0)}{r_{13} \cdot (X - X_0) + r_{23} \cdot (Y - Y_0) + r_{33} \cdot (Z - Z_0)}$$

In computer vision, those parameters are solved in the image coordinates system (\mathbf{x}, \mathbf{y} and \mathbf{z} along the viewing direction) rather than world coordinates. In our implementation, camera position and orientation are derived from the translation and rotation matrix computed with OpenCV (<http://opencv.willowgarage.com>). Once pose is known, it can be used in the projective function to compute image coordinates of world objects.

DEM projection in the image

The DEM is projected in the image using Equation 1. This step selects the 3D raster cells which are located in the camera viewing pyramid. In order to populate each pixel with the 3D coordinates of the closest world point, potential 3D raster cells are triangulated with the scientific tool for python (scipy: <http://www.scipy.org/>). Triangles are projected in the image plane and rasterized in pixel units. The pixels contain a depth value (z-buffer technique [11]), to ensure that the closest world point is the one being attributed at the end of the process. At this time, each pixel consists of a pixel coordinate, an intensity value, a z-depth and a 3D world coordinate.

The pixels on the image boundary are merged to draw the footprint polygon. During ortho-rectification, pixels are mapped on the XY plane and triangulated. As the triangulated pixels are not regularly distributed, intensity values are interpolated on a regular grid, thus providing the orthophoto. Inversely, the assignation of an ortho-image intensity value to the image pixels generates a synthetic image. The ortho-images used for the synthetic image are generated at different scales and slightly blurred in order to generate the most natural synthetic image.

Results

Two images (Figures 3a and 3b) of the Aletsch glacier taken between 1890 and 1900 are downloaded from the **library of congress** (<http://www.loc.gov/pictures/>). Those pictures belong to a collection called “View of Switzerland”. They are taken from the top of the Eggishorn.

FIGURE 3

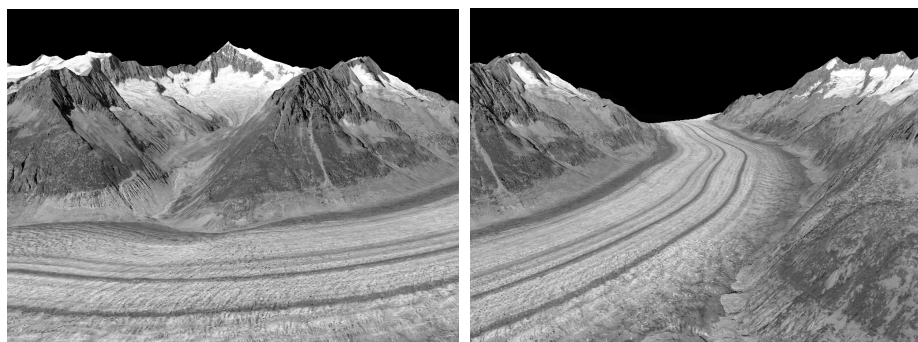


A

B

Ground Control Points are found in the image and in the up-to-date ortho-image. Fortunately, those landscape areas do not change much over 100 years and GCP are stable. However, a trained user can spent between 30 and 60 minutes on each picture. The accuracy and the number of the GCP (7 and 8 respectively) are not sufficient to fully calibrate the camera. The principal point is assumed to be at the image center. The focal that minimizes the reprojection error of each 3D point on the image plane is used. This focal generates a camera location close to Eggishorn peak, which is expected to be the photographer location. The mean reprojection errors are 6.2 and 3.3 pixels respectively.

The camera pose computed is used to project the DEM in the image plane. Synthetic images (Figures 4a and b) are generated, where we can appreciate the changes occurring in the last century. In Figure 4a, the lateral glacier size is greatly reduced and do not reach the main glacier anymore. In Figure 4b, in the lower right corner, the glacier retreats and its height decreases. Note that the mountains in the background are not represented because they are out of the DEM considered for the study area.

FIGURE 4

A

B

These ortho-rectified-images are used to evaluate the right bank change between 1900 and 2008. The glacier bound is drawn on the ortho-rectified ancient image and on the current ortho-image. Figure 5 expresses those differences with 4 profiles. The 1900 glacier level is drawn with a green line, the 2008 one is in red. Table 1 summarizes the difference found in the ancient and current glacier state.

TABLE 1

Profil	Horizontal difference [m]	Vertical difference [m]
A	311	82
B	191	120
C	202	83
D	130	44

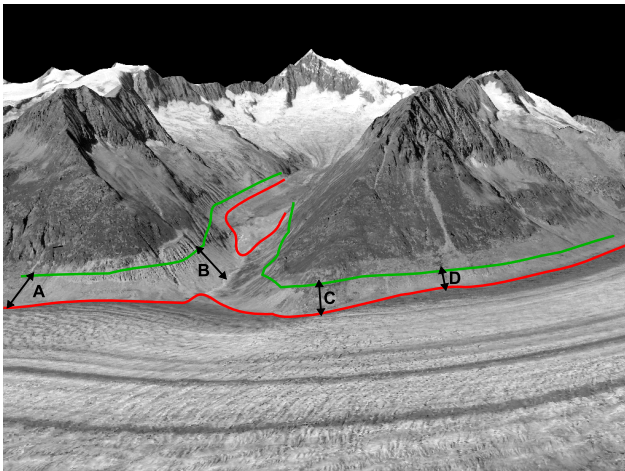


FIGURE 5

Conclusion

Ancient pictures are often used to assess glacier reduction or any environmental change during the last century. The repeated photography technique [12], [13] is widely used to compare those images with a recent image taken from the same location. However, with this process the comparison cannot be quantified. Monoplotting softwares allow such comparison without going to the field, find the right location and take a new picture, which is very challenging for some pictures. Moreover, monoplotters assign a world coordinate to each pixel, thus allowing the quantification of the change.

The python implementation proves to be fast and robust. Scipy and Numpy libraries bring Python close to common mathematic programming languages (Matlab, Octave) that scientist are used to work with. However, since Python is also a general programming language, it is easy to link to open software as openCV. OpenCV is a very active project in computer vision and cannot be substituted by any commercial one. Moreover, Python is widely used in the GIS community and its choice will facilitate interaction. Our own 3D rendering software was needed to support extended and very specific functions. Further development will investigate faster processing on Graphics Processing Units (GPU).

[1] BJORK, A. A., KJAER, K. H., KORSGAARD, N. J., AND KHAN, S. An aerial view of 80 years of climate-related glacier fluctuations in southeast Greenland. *Nature Geoscience* (2012), 427–432.

[2] BOZZINI, C., CONEDARA, M., AND KREBS, P. A new tool for obtaining cartographic georeferenced data from single oblique photos. *Proceedings of the 23rd International CIPA Symposium* (Prague, Czech Republic, 2011).

[3] MITSHITA, E. A., MACHADO, A. L., HABIB, A. F. AND GONCALVES, G. 3D monocular restitution applied to small format digital airphoto and laser scanner data. *Proceedings of Commission III, ISPRS Congress* (Istanbul, Turkey, 2004).

- [4] JAUREGUI, M., VILCHEZ, J., AND CHACON, L., A procedure for map updating using digital mono-plotting. *Computer & Geosciences* (2002), 513-523.
- [5] DOYSTER, Y. AND HALL, J. K., FORTRAN programs for coordinate resection using an oblique photograph and high-resolution DTM. *Computer & Geosciences* (1995), 895-905.
- [6] MAKAROVIC, B. Data base updating by digital mono-plotting. *Environmental assessment and resource management* (1983), 433-442.
- [7] MAKAROVIC, B. Digital mono-plotters. *ITC Journal* (1973), 583-599.
- [8] CORRIPIO, J. G., Snow surface albedo estimation using terrestrial photography. *International Journal of Remote Sensing* (2004), 5705 – 5729.
- [9] ASCHENWALD, J., LEICHTER, K., TASSER, E. AND TAPPEINER, U. Spatio-temporal landscape analysis in mountainous terrain by means of small format photography: a methodological approach. *Geoscience and Remote Sensing* (2001), 885-893.
- [10] FLUEHLER, M., NIEDEROEST, J., AKCA, D., Development of an educational software system for the digital monoplotting. Proceedings of *ISPRS on Tools and Techniques for E-Learning*. (Potsdam, Germany, 2005).
- [11] COOK, R. L., CARPENTER, L., AND CATMULL, E. The Reyes image rendering architecture. *ACM SIGGRAPH Computer Graphics* (1987), 95-102.
- [12] ROUSH, W., MUNROE, J. S., AND FAGRE, D. B. Development of a spatial analysis method using ground-based repeat photography to detect changes in the alpine treeline ecotone, Glacier National Park, Montana, USA. *Arctic, Antarctic, and Alpine Research* (2007), 297-308.
- [13] KULL, C. Historical landscape repeat photography as a tool for land use change research. *Norwegian Journal of Geography* (2005), 253-268.

Open tools and methods for large scale segmentation of Very High Resolution satellite images

Authors

- Julien Michel, CNES (DCT/SI/AP), *France*
- Manuel Grizonnet, CNES (DCT/SI/AP), *France*
- Arnaud Jaen, CS, *France*
- Sébastien Harasse, CS, *France*
- Luc Hermitte, CS, *France*
- Jonathan Guinet, CS, *France*
- Julien Malik, CS, *France*
- Mickaël Savinaud, CS, *France*

KEYWORDS : Open-source software, segmentation, remote sensing images, GIS

1. Introduction

With the increase of the spatial resolution of satellite images, analysis techniques such as Object Based Image Analysis or Spatial Reasoning [4] have become widely studied and used. Because they use objects rather than pixels as their primitives, these methods are very well adapted to represent and extract the information contained in very high resolution imagery (VHR). Moreover, reasoning on objects is often supported by very sound theories. Yet one of their severe weakness is the process of obtaining the objects themselves: segmentation is widely used as a pre-processing step for these techniques, and it is well known that the task of segmenting all categories of object of interest, across a large very high resolution scene and with a controlled quality is a difficult task for which no method has reached a sufficient level of performance to be considered as operational.

Even if we leave aside the question of segmentation quality and consider that we have a method performing reasonably well on our data and objects of interest, the task of scaling up segmentation to real very high resolution data is itself challenging. First, we can not load the whole data into memory, and there is a need for on the flow processing which does not cope well with traditional segmentation algorithms [5]. Second, the result of the segmentation process itself is difficult to represent and manipulate efficiently.

There are, to our best knowledge, few open source softwares able to overcome these issues. In the frame of the development of the Orfeo ToolBox [7], we therefore initiated some work to provide software components for this purpose.

2. Data representation and conversion

In this section, we review three standard structures to store segmentation results, and explain why we selected the third one for our framework.

The most common way of representing segmentation results from an image processing perspective is to derive a raster where each pixel contains the unique label of the segment it belongs to. Such a raster is also known as label image. Yet this representation has numerous drawbacks. First, accessing all pixels of a given segment identified by its label requires parsing the whole label image. Second, storing large segmentation results with billions of segments might require a high number of bits to represent the label, thus increasing dramatically the size of the output. Last, the constraint of label uniqueness is very strong, yet not very useful: it is required since label images only provides an implicit description of segments, based on neighbouring pixels with same labels. This representation is of limited interest for our purpose.

A second representation of segmented image which is further from image processing is the map of label objects [3]. A segment is represented in Run Length Encoding (RLE), and all segments are indexed by a unique label in a map. The RLE representation is more compact, the map allows fast access to a segment given its label, and it also allows to store attributes related to a segment, which is a first step toward Object Based Image Analysis. The main drawback of this representation is that there are no standard file format

able to store it on a file system, and it requires to convert either to or from raster or vector representation at each read or write operation. Moreover, this representation is neither really raster nor really vector. Therefore, operations like walking pixels in the segment or following its contour are more complex.

Neither of these two structures can fit our needs: the first one is highly inefficient, while the second one lacks of existing file formats. We chose to store our segmentation results as boundaries in a vector structure and file format. This has numerous advantages. First, it overcomes the issue of scaling up to large image segmentation: a collection of vector objects, i.e. polygons or multi-polygons, is a compact representation which will grow linearly with the size of the image, and does not need an explicit unique indexing label. Second, there are several file formats and even databases to represent this kind of vector data, especially in the GIS world. Last, using such formats guarantees a full interoperability between the segmentation tools and most GIS software, which is desirable.

As described in the next section, we intend to derive a generic framework for large scale remote sensing images segmentation, thus we need to be compatible with the raster output of most segmentation algorithms, which usually produce label image. For this purpose, we need a component for exact conversion between the raster representation and the vector one. We take advantage of the polygonization and rasterization algorithms provided by GDAL [2]. To handle on the flow input and output to vector data files and databases, we then use the full extent of OGR functions [6]. This abstraction layer allows our tool to address simple file formats like ESRI shapefile or complex databases like PostGIS seamlessly.

3. A generic framework for large scale segmentation

We define a segmentation algorithm in the Orfeo ToolBox as a filter that accepts a remote sensing image as input and produces a label image as output. This filter is not supposed to have streaming (on the flow) capabilities, it is allowed to require the whole input image to produce its output, and it can make an internal use of multi-threading. Given such an algorithm, and a large image to segment (for instance a basic Pleiades [8] image has an extent

of 40 000 by 40 000 pixels), our framework works as follows:

- Derive a tiling scheme according to the amount of memory available on the computer and additional information such as file format efficiency or user defined parameters
- For each tile of the tiling scheme :
 - a. Load the corresponding image part into memory
 - b. Segment the image extract with the segmentation algorithm
 - c. Polygonize the result using a filter based on GDAL capabilities
 - d. Dump the polygons to file system or database through OGR abstraction

The strength of this framework is that it will naturally scale up: larger images will take more time to process and more space on file system to store, but the system will never run out of the most limited resource on most hardware, which is memory. The most time consuming operation is the segmentation, and this is where we allow for parallel processing depending on the segmentation algorithm implementation. For instance, we implemented a multi-threaded version of the Mean shift algorithm [1].

A key feature is that we make few assumptions on the segmentation algorithm, and that our framework will be able to perform with any segmentation algorithm implemented under these assumptions. The same code already runs with a basic connected components algorithm, two different implementations of the Mean shift algorithm, and Watershed.

Of course, it has some drawbacks, the first one being that region lying on tiles borders will be artificially split by the tiling scheme. We address these issues in the following section.

4. Pre and post-processing to enhance usability

In order to get useful results from this large scale segmentation, we need to address several issues. The first one issue is the splitting of the regions on the border of tiles. The naive solution we implemented is a simple stitching rule: we post-process the vector data by looking for neighbour polygons lying on each side of a tile border and merge them if their contact surface is large enough. More sophisticated techniques might be derived in the future.

The second problem is that the polygons reflect exactly the shape of the

segment, and contain a large amount of vertices. This leads to very heavy vector files, sometimes even heavier than the corresponding raster image would be. To deal with this issue, we added pre-processing as well as post-processing. First, we added an input mask image to avoid segmenting unwanted regions, like no-data pixels, clouds, or vegetation (if vegetation is not desired). We also added a rule to remove very small segments, which are more likely to correspond to segmentation noise than to real objects (this leads to holes in the segmentation canvas). Last, we again used the capabilities of OGR to perform a geometry simplification algorithm, which simplify polygons by removing vertices according to a given tolerance.

These additional processing allows to produce a cleaner vector segmentation output which can then be used in a GIS software. The figure below shows a Mean-Shift segmentation of an extract of a Pleiades image over Las Vegas displayed in QGIS (copyright CNES).

5. Conclusion and Future Work

This framework is only a first step toward providing an open source large scale OBIA and spatial reasoning framework. Future developments will

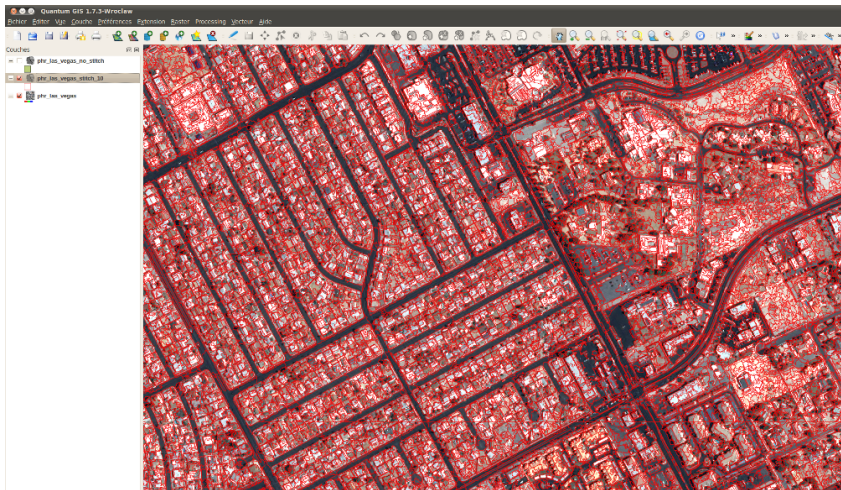


FIGURE 1

include more stitching and tiling strategies to enhance the segmentation performance at tiles borders, and the computation of attributes along with polygons, such as statistics on radiometry from an image, or shape attributes. These attributes could then be used for reasoning or classification at the object level. While there is still a lot missing, we hope to provide a comprehensive environment for object-based techniques for VHR images analysis through the ongoing efforts to integrate Orfeo ToolBox within Quantum GIS [9], a open source GIS software, via Sextante [10], and bridge the gap between remote sensing and GIS.

[1] COMANICIU, D., MEER, P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, pp. 603–619, 2002.

[2] Gdal project homepage. Retrieved September 6, 2012 from <http://www.gdal.org/>

[3] LEHMANN, G. Label object representation and manipulation with itk. *Insight Journal*, pp. 1–34, 2008.

[4] INGLADA, J., MICHEL, J. Qualitative spatial reasoning for high-resolution remote sensing image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 2, pp. 599–612, 2009.

[5] SHI, J., MALIK, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[6] OGR project homepage. Retrieved September 6, 2012 from <http://www.gdal.org/ogr/>

[7] Orfeo ToolBox project homepage. Retrieved September 6, 2012 from <http://www.orfeo-toolbox.org/otb/>

[8] CNES, Pleiades project homepage. Retrieved September 6, 2012 from <http://smc.cnes.fr/PLEIADES/Fr/>

[9] Quantum GIS project homepage. Retrieved September 6, 2012 from <http://www.qgis.org/>

[10] Sextante project homepage. Retrieved September 6, 2012 from <http://www.sextantegis.com/>

Utilization of the Scythe C++ open source library for statistical geocomputation in livestock landscape genomics

Authors

- **Sylvie Stucki**, Ecole polytechnique fédérale de Lausanne, *Switzerland*
- **Saif Agha**, Ain Sham University, *Egypt*
- **Menghua Li**, MTT Agrifood Research Finland, *Finland*
- **Stéphane Joost**, Ecole polytechnique fédérale de Lausanne, *Switzerland*

KEYWORDS : Landscape genomics, Scythe C++ open source library, spatial analysis, natural selection

Introduction

Understanding genetic basis of adaptation is of key importance, especially to insure food supply for present and future populations. Cultivated plants and livestock animals are subjected to high selection pressure due to climate change, increasing food needs and intensive methods in agriculture. Widespread use of artificial selection emphasizes short-term productivity. Native breeds, known to be robust and well adapted, are neglected and some of them are disappearing [1]. In this context, genetic diversity of domestic species has to be assessed and protected [2].

Since adaptation processes are intrinsically linked to landscape, spatial analysis has a role to play in explaining how environmental features shape genetic pool. A research field named landscape genetics combines spatial analysis, population genomics and molecular ecology to provide information about the

interaction between landscape features and micro-evolutionary processes [3]. This discipline focuses on the spatial dimension of genetic information given by the location of organisms under study. Based on the spatial coincidence concept [4], it aims at explaining spatial genetic patterns by using landscape variables [5, 6]. It developed quickly and yielded to new perspectives [7].

Landscape genomics was first mentioned as the possibility to perform direct association studies between genome and environment [8, 9]. This field is included in landscape genetics, but refers more specifically to the use of the forthcoming large amount of genetic data due to high-throughput sequencing [7]. This combination of bioinformatics, genomics, spatial statistics and landscape ecology aims to predict the frequency of a genetic marker on the basis of environmental parameters by building multiple association models. It allows the detection of genes under natural selection, but also to discover which landscape or climatic traits are shaping adaptation [10, 11]. Recent developments in high-throughput sequencing enabled the detection of several millions variable sites across the whole genome [12]. Simultaneous advances in biotechnology and environmental data availability can combine with current computation capacity to allow landscape genomics to develop its full potential.

Although association studies may be conducted with standard statistical software, processing such large amount of genetic data provided by high-throughput sequencing technologies is computationally intensive. Therefore we decided develop an open source toolbox dedicated to landscape genomics. The implementation was made easier by using Scythe, an open source C++ library for statistical computation [13]. It supports matrix calculation, provides various statistical functions and allows faster computations than interpreted languages. We considered using LAPACK [14] or ALGLIB [15] instead. These libraries are also open source and LAPACK is widely used for scientific computing. We chose Scythe for its intuitive programming interface and its header-only structure that ease packaging. Scythe suited our need for improved performance compared to MatSAM, the initial software tool developed with Matlab [16]. The program is written in C++, is stand-alone and multi-platforms. In order to ease results handling, it sorts the models according to their statistical significance and can discard on the fly irrelevant associations.

Here, the case study we present illustrates the application of this development to 37 domestic sheep breeds, in the genome of which we looked for signatures of natural selection with the help of a combination of genetic markers and topo-climatic variables.

Data

Genetic data is made of specific DNA sequences called genetic markers. These markers act as reference points to follow the transmission of chromosome segments from one generation to another. We used Single Nucleotide Polymorphisms (SNPs) - a particular type of genetic marker - collected in the context of the Sheep HapMap project [17]. These SNPs are precisely located (a single location is named a locus) in the genome of the individuals sampled and can be present (1) or not (0) on one or on the two homologous chromosomes (the same location on one of the two homologous chromosomes is named an allele).

The data matrix was made of 1'479 lines (the number of individuals selected from 37 sheep breeds), and 147'102 columns (binary variables) resulting from the recoding of the original 50'000 SNPs markers (the number of genomic regions characterized). Indeed, several allele combinations are possible at each locus. The matrix was completed by 105 columns containing environmental variables characterizing the habitat of the 1'479 animals analysed. The geographic location of the latter was provided by HapMap contributors [18, 19]. Environmental information consisted in monthly and yearly means

Variable	Description
altitude	altitudes recorded on the field and completed with SRTM3 when required
dtr	monthly values of mean diurnal temperature range in °C
frs	monthly values of number of days with ground-frost
pr	monthly values of precipitations in mm/month
prcv	monthly values of the coefficient of variation of monthly precipitation in percent
tmp	monthly values of mean temperature in °C
rdo	monthly values of wet-days (number of days with < 0.1 mm rain per month)
reh	monthly values of relative humidity in percent
sun	monthly values of percent of maximum possible sunshine (percent of day length)

TABLE 1

Name of the topo-climatic variables, their description and abbreviation

of eight climatic variables provided by the Climatic Research Unit (CRU) [20] (Table 1). Monthly variables were considered because in sheep several management and production systems based on lambing periods (or seasons) are used in different breeds (e.g. autumn lamb production, winter lambing, or spring lamb production) [9]. In addition, altitude was retrieved from the Shuttle Radar Topography Mission (SRTM30) [21]. Thus 105 environmental parameters were gathered for analysis.

Method

Genetic material was gathered by sampling individuals in the field, and the record of their geographic locations with GPS devices permitted to retrieve the environmental features characterizing their habitats.

To calculate the relationship between genetic markers and environmental variables, and since the dependent variable is binary (presence or absence of a given marker), we computed many parallel logistic regressions to model their frequencies by maximizing the log-likelihood function [22]. Robustness of association was tested by computing the log-likelihood ratio G and the Wald statistics [9]. We chose a conservative filter in order to focus on the most significant relationships among the large amount of models to process. The program computed all univariate and multivariate [22] models and ranked them according to their Wald score.

To validate the results produced by the spatial analysis method, the data set was also analysed with the help of a program based on a theoretical population genomics approach. The latter relies on the principle that in a given population, all genomic regions have a comparable differentiation index (F-Statistics based on frequency analysis) because they share identical demographic histories. Only regions subject to natural selection have their F-Statistics modified [23]. Here we used Lositan (<http://popgen.eu/soft/lositan/>), an open source Java web start application [24].

Results

Four genomic regions were detected to possibly be under natural selection with a Wald score threshold of 385. These markers are GG_OARX_63571789.1, AA_OARX_63571789.1, AA_s43015.1 and AA_OARX_59578440.1. They are mainly associated with summer weather conditions: sunshine, number of rainy days and mean diurnal temperature range.

Table 2 presents the number of significant models and markers regarding to the threshold. Table 3 shows the 29 most significant models, their statistical scores and the results provided by Lositan, which shows the probability of each marker to be under selection.

– log(p-value)	Wald score threshold	Number of models	Number of markers
97	473.19	1	1
95	464.00	1	1
90	441.03	8	2
85	418.05	16	2
80	395.08	25	4
75	372.12	49	7
70	349.15	119	20
65	326.20	249	40
60	303.24	521	70
55	280.30	1020	115
50	257.35	1962	227
45	234.42	3812	447
40	211.50	7664	899
35	188.59	15632	1634
30	165.69	31568	2929
25	142.81	63041	5115
20	119.95	124677	8737
15	97.14	248932	14836
10	74.37	503415	25728
5	51.70	1048972	46139
3	42.67	1425159	58290

TABLE 2

Change in the number of significant models regarding to the threshold

Genotype	Env_1	Gscore	WaldScore	Type of correlation	P
GG_OA RX _63571789.1	SUN-September	960.48	474.39	-	0.995558
	SUN-July	967.51	452.89	-	0.995558
	SUN-August	955.00	449.16	-	0.995558
	DTR-September	879.24	447.13	-	0.995558
	RDO-August	841.10	439.61	+	0.995558
	SUN_Yearly_Mean	729.57	439.10	-	0.995558
	RDO-July	835.66	425.72	+	0.995558
	SUN-June	806.74	423.82	-	0.995558
	SUN-October	673.07	421.55	-	0.995558
	DTR-August	894.66	412.06	-	0.995558
	PRCV-October	691.88	404.25	-	0.995558
	RDO-September	813.26	403.38	+	0.995558
	DTR_Yearly_Mean	609.61	396.53	-	0.995558
	DTR-October	626.63	396.39	-	0.995558
DTR-July	934.35	385.04	-	0.995558	
AA_OA RX _63571789.1	SUN-August	796.92	458.60	+	0.995558
	RDO-August	803.47	458.28	-	0.995558
	SUN-July	823.13	455.19	+	0.995558
	SUN-September	778.69	444.24	+	0.995558
	RDO-July	828.87	440.00	-	0.995558
	DTR-September	882.40	437.36	+	0.995558
	SUN-June	643.30	420.04	+	0.995558
	DTR-August	936.11	402.05	+	0.995558
	DTR-July	1005.49	395.98	+	0.995558
DTR-June	794.20	385.18	+	0.995558	
AA_s43015.1	SUN-October	687.88	401.91	-	0.994685
	RDO-January	656.77	390.10	+	0.994685
AA_OA RX _59578440.1	RDO-August	601.51	397.51	+	0.999846
	RDO-July	635.21	385.45	+	0.999846

TABLE 3

List of the most significant models constituted by environmental variables associated with genotypes. The type of correlation is positive («+») when the frequency of the genotype increases with the value of the environmental parameter. The last column (P) is the probability computed by Lositan that the simulated differentiation index (F_{ST}) is smaller than the observed one. The high P means that the probability is high that the locus is under selection

Discussion

Results show that genetic markers can be involved in several significant associations. Such multiple matches are common since local adaptation can be driven by a combination of different environmental variables. In our results, this effect is increased because of the temporal resolution of the

environmental data. Consecutive months of the same variable are expected to be mutually correlated, thus a genetic trait associated with this environmental characteristic is likely to show several significant associations. See for instance how the two markers involved in the most significant models are associated with the percentage of maximum sunshine in summer (Table 3).

The four markers displayed in Table 3 are also detected to be outliers by Lositan. The two most significant markers (GG_OARX_63571789.1 and AA_OARX_63571789.1) are located at the same locus. They tend to show strong associations with common environmental variables in a symmetrical way: the variant GG is frequent where summer is wet, and rare where summer is hot and sunny. The AA variant shows the opposite behaviour. The third possible variant for this locus (GA_OARX_63571789.1) is not influenced by summer conditions but is associated with dry winters and rainy springs, but with a lower significance level (Wald score < 70).

Recent developments in sequencing technologies along with the gradually increasing availability of environmental data and of computation capacity have brought the emerging field of landscape genomics to a productive stage. This method is complementary to standard population genetics approaches in detecting signatures of selection while identifying environmental variables possibly shaping local adaptation. Indeed, the knowledge of the nature of the selection pressure may facilitate the emergence of working hypotheses about the role of the regions of the genome which are linked to the analysed markers.

To this end, we applied the landscape genomics approach with a program developed on the basis of an open source C++ library for statistical computation named Scythe, much faster than a previous version written in MATLAB. It permitted to efficiently identify 4 genetic markers significantly associated with specific environmental variables exerting a selection pressure on the genome of the 37 sheep breeds under study.

Acknowledgments

The ovine SNP50 HapMap dataset was provided by the International Sheep Genomics Consortium (www.sheepmap.org), in agreement with the ISGC Terms of Access.

S. Agha and M. Li were funded by the European Science Foundation (ESF), in the context of the Genomic Resources Research Networking Programme (Exchange grants n° 3271 and 3259).

The authors would like to thank Marco Milanesi and Paolo Ajmone-Marsan who helped with the elaboration of the dataset and provided useful comments.

[1] TABERLET, P., AND VALENTINI, A., AND REZAEI, H. R., ET AL. Are cattle, sheep, and goats endangered species? *Molecular Ecology* 17 (2008), 275–284.

[2] FOOD AND AGRICULTURE ORGANISATION. *Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration*. Rome, 2007.

[3] MANEL, S., AND SCHWARTZ, M. K., AND LUIKART, G., ET AL. Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology and evolution* 18 (2003), 189–197.

[4] GOODCHILD, M. F. Geographic Information Systems and spatial analysis in the social sciences. In *Anthropology, space, and Geographic Information Systems*. ALDENDERFER, M., AND MASCHNER, H. D. G., Eds. Oxford University Press, New York, 1996, 214–250.

[5] STORFER, A., AND MURPHY, M. A., AND EVANS, J. S., ET AL. Putting the 'landscape' in landscape genetics. *Heredity* 98 (2007), 128–142.

[6] JOOST, S., AND COLLI, L., AND BARET, P. V. Integrating geo-referenced multiscale and multidisciplinary data for the management of biodiversity in livestock genetic resources. *Animal Genetics* 41 (2010), 47–63.

- [7] MANEL, S., AND JOOST, S., AND EPPERSON, B. K., ET AL. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* 19 (2010), 3760–3772.
- [8] LUIKART, G., AND ENGLAND, P. R., AND TALLMON, D., ET AL. The power of population genomics: from genotyping to genome typing. *Trends in ecology and evolution* 18 (2003), 189–197.
- [9] JOOST, S., AND BONIN, A., AND BRUFORD, M. W., ET AL. A Spatial Analysis Method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* 18 (2007), 3955–3969.
- [10] BEAUMONT, M. A., AND NICHOLS, R. A. Evaluating loci for use in the genetic analysis of population structure. *Proceedings Royal Society London B* 263 (1996), 1619–1626.
- [11] SCHWARTZ, M. K., AND LUIKART, G., AND MCKELVEY, K. S., ET AL. Landscape Genomics: A Brief Perspective. In *Spatial Complexity, Informatics and Wildlife Conservation*. CUSHMAN, S. A., AND HUETTMANN, F., Eds. Springer, New York, 2009, 165–174.
- [12] ALTSHULER, D. L., AND DURBIN, R. M., AND ABECASIS, G. R., ET AL. A map of human genome variation from population-scale sequencing. *Nature* 467 (2010), 1061–1073.
- [13] PEMSTEIN, D., AND QUINN, K. M., AND MARTIN, A. D. The Scythe Statistical Library: An Open Source C++ Library for Statistical Computation. *Journal of Statistical Software* 42 (2011), 1–26.
- [14] ANDERSON, E., AND BAI, Z., AND BISCHOF, C., ET AL. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, 1999.
- [15] BOCHKANOV, S., AND BYSTRITSKY, V. 2012. ALGLIB. Retrieved August 23, 2012 from <http://www.alglib.net>
- [16] JOOST, S., AND KALBERMATTEN, M., AND BONIN, A. A Spatial Analysis Method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular Ecology Resources* 8 (2008), 957–960.
- [17] INTERNATIONAL SHEEP GENOMICS CONSORTIUM. 2010. *Sheep HapMap and Animal Resources*. Retrieved June 12, 2012 from www.sheepmap.org/hapmap.php

[18] AGHA, S. 2011. Landscape genomics and FST approaches to reveal SNPs under selection in Cattle. *European Science Foundation*. Retrieved June 10, 2012 from <http://www.esf.org/activities/research-networking-programmes/life-earth-and-environmental-sciences-lesc/current-esf-research-networking-programmes-in-life-earth-and-environmental-sciences/advances-in-farm-animal-genomic-resources-genomic-resources/grants/grant-awards.html>

[19] LI, M.-H. 2010. Application of modern high-throughput molecular technology and a multidisciplinary approach in characterizing the sheep genetic resources in the North. *European Science Foundation*. Retrieved June 10, 2012 from <http://www.esf.org/activities/research-networking-programmes/life-earth-and-environmental-sciences-lesc/current-esf-research-networking-programmes-in-life-earth-and-environmental-sciences/advances-in-farm-animal-genomic-resources-genomic-resources/grants/grant-awards.html>

[20] NEW, M., AND LISTER, D., AND HULME, M., ET AL. A high-resolution data set of surface climate over global land areas. *Climate Research* 21 (2002), 1–25.

[21] NATIONAL AERONAUTICS AND SPACE ADMINISTRATION. 2009. *The Shuttle Radar Topography Mission*. Retrieved June 10, 2012 from <http://www2.jpl.nasa.gov/srtm/>

[22] DOBSON, A. J., AND BARNETT, A. G. *An Introduction to Generalized Linear Models*. Chapman & Hall, New York, 2008.

[23] PEREZ-FIGUEROA, A., AND GARCIA-PEREIRA, M. J., AND SAURA, M., ET AL. Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology* 23 (2010), 2267–2276.

[24] ANTAO, T., AND LOPES, A., AND LOPES, R., ET AL. LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9 (2008), 323.

MCDA-GIS integration: an application in GRASS GIS 6.4

Authors

- **Gianluca Massei**, University of Perugia, *Italy*
- **Lucia Rocchi**, University of Perugia, *Italy*
- **Luisa Paolotti**, University of Perugia, *Italy*
- **Salvatore Greco**, University of Catania, *Italy*
- **Antonio Boggia**, University of Perugia, *Italy*

KEYWORDS : GIS, Multicriteria Analysis, Spatial Decision Support Systems, Dominance-based Rough Set Approach

To face a spatial decision problem means to apply both Geographical Information Systems (GIS) and a Multi-Criteria Decision Analysis (MCDA) approach [15]. GIS and MCDA are two well defined research areas, but both can benefit one from each other [15]. MCDA methods are a basic tool in the field of environmental valuation, but they cannot easily take into account the spatial dimension. We can define MCDA as “a decision-aid and a mathematical tool allowing the comparison of different alternatives or scenarios according to many criteria, often contradictory, in order to guide the decision maker(s) towards a judicious choice” [19]. The typical MCDA problem is represented as a set of m alternatives to be valued in terms of a vector of n criteria. Sometimes, the Decision Maker’s (DM) subjective preferences are represented by the weight vector w . Application of MCDA methodologies give as a solution a recommendation to classify, rank or choose efficient alternatives [19]. As a result of distinctive assumptions, the different MCDA methodologies differ in many aspects, related to the aggregation procedure, which aims at synthesizing the often conflicting evaluations on criteria, and to the exploitation procedure, which aims at using the result

from the aggregation procedure, in order to define a proper recommendation. The conventional MCDA approach assumes spatial homogeneity of alternatives within the case study area, although this is often unrealistic [14]. If alternatives have a geographical classification, ordering or choosing operations have to depend also on their spatial arrangement [14]. On the other hand, GIS provide great capabilities of data acquisition, storage, manipulation and analysis. However, in case of a value structure analysis, this capability is lower [14]. Due to these reasons, in the last twenty years several researchers have paid attention to MCDA-GIS integration and to the development of Spatial Decision Support Systems (SDSS) [2, 3, 10, 11, 14, 15].

MCDA-GIS integration consists in combining value judgments and geographical data, but also their transformation and elaboration [15]. The evaluation procedure needs both the MCDA framework and the GIS possibilities. According to several authors [2, 12], it is possible to classify MCDA-GIS integration in a three-level system. The basic step is the MCDA-GIS indirect integration: MCDA and GIS models are separated, and connected through an intermediate connection system. In the Built-in MCDA-GIS models, the multicriteria part is a component integrated into the GIS system, but it is independent from a logical and functional point of view. Only with the complete, or full, MCDA-GIS integration [15] the two systems use the same interface and the same database. The MCDA model is activated inside the GIS software just like any other analysis function.

We developed an example of a complete integration, implementing five modules, based on five different Multi-Criteria Decision Analysis methodologies, in GRASS GIS 6.4 svn. The methods used in the modules are Electre [18], Fuzzy set [21, 22, 23], Regime [9, 17], Analytic Hierarchy Process [20] and Dominance-based Rough Set Approach [6, 8]. The syntax of each module, based on the GRASS GIS one, is the following: ***r.mcda.[algorithm]***. Prefix ***r*** means “raster,» i.e. the type of data elaborated; ***mcda*** is the name of the package, while ***[algorithm]*** is the name of the MCDA method applied. For instance, the module corresponding to the method Electre is named ***r.mcda.electre***, and so on.

We chose GRASS GIS for this application because it is an advanced and well known open source GIS software, used for geospatial data management and analysis, image processing, graphics/maps production, spatial modeling and visualization. Since its first release in 1982 [5], GRASS GIS has been increasingly used by academic and commercial settings all around the world, as well as by many governmental agencies and environmental consulting companies, for the wide range of possible applications [4, 13, 16]. GRASS GIS is a leading open source GIS software. Moreover, it is written in C language, and it is possible, thanks to its libraries and GPL license, to develop easily new modules [16, 5].

We tested the possibilities of our tool in an application to a geological instability management problem. The objective of the application is to identify several zones where realize agro-environmental measures, for territory maintenance and for reducing hydro geological risk. The case study is the area of the hydrographic basin of Menotre (Umbria, Central Italy), a small river, which is extended over 24.000 ha. The basin has an altitude between 257 and 1478 m above sea level. The upper trait is characterized by natural or rural ground, while in the valley it is quite completely urbanized. The basin is largely interested by geological instability, which is not uniform because of the geological characteristics of the ground, the presence or absence of vegetation, the level of urbanization and the type of agricultural activities.

The module used in the case study is the *r.mcda.roughset*, which represents the first implementation of DRSA in a spatial setting. Several algorithms are available for implementing DRSA. One of the most well-known and widely used is the iterative algorithm DOMLEM [7], which was found to be particularly accurate with large datasets. For this reason, DOMLEM seems particularly suitable for GIS applications, and we chose to use it in the *r.mcda.roughset* module.

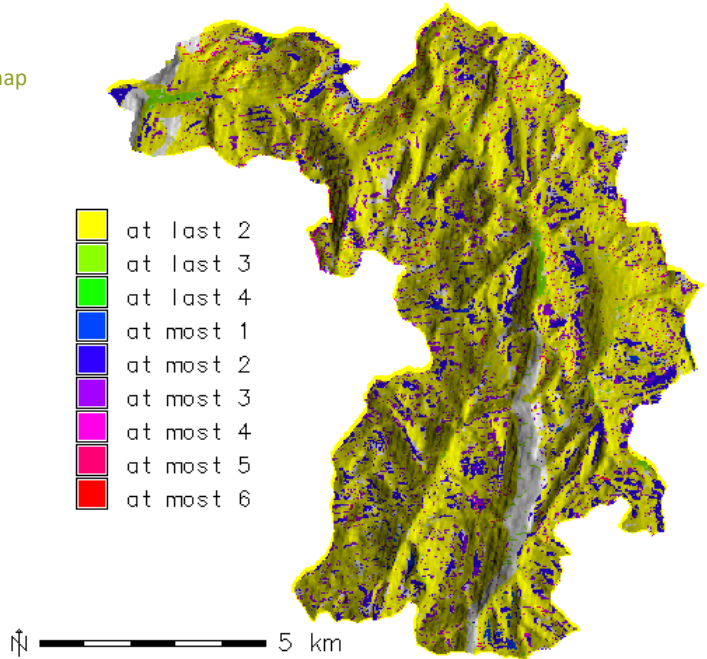
DRSA is different from the majority of other MCDA methods, because it does not involve weights to represent the importance of the considered criteria. In the classical multicriteria methods, to fix the weights of the considered

criteria, the DM is often asked to answer questions that require a great cognitive effort, which reduces the reliability of the preference information so obtained. Thus, the non-consideration of weights constitutes a strong point of the DRSA. In the DRSA, the request for more or less direct information on the weights to the DM is replaced by the request of exemplary decisions, in terms of classification of some minimal units, well known to the DM, in predefined classes. As a result of the application of the DRSA methodology, the DM is supplied with some “*if..., then...*” rules, explaining the exemplary decisions. Those rules are then used to classify all the minimal units of the GIS. One advantage of the integration of DRSA with GIS is that the exemplary decisions require a much lower cognitive burden with respect to the input required by other GIS methods (e.g., weights of criteria). Additionally, the decision rules and the results of their applications are easily understandable because they are expressed in a natural and simple language. There are three types of rules: certain, possible and ambiguous. In the module, the rules are derived from a raster map, which includes a thematic key essential in the analysis. There are two module outputs: a raster map and a text file. The text file is compatible with 4emka2, Jamm [6] and JMAF [1], which are three software packages developed by the Poznan University of Technology.

The inputs required in the string are the name of the criteria raster map (***criteria=name[,name,...]***), the preferences in terms of gain and cost (***preferences=character***), the name of the decision raster map (***decision=name***), the name of the classified output raster map (***outputMap=string***) and the name of the output txt files (***outputTxt=name***). The criteria raster maps we used are: the tendency toward instability, the Normalized Difference Vegetation Index (NDVI), the slope, the LS factor from the RUSLE model and the use of the ground. As decision raster map we used the frequency of surface landslide in the last ten years. We supposed that the probability to have another landslide will be higher in case of previous events. So a low value for this theme corresponds to a low need for management measures.

FIGURE 1

Ordinal decision map



Through the elaboration of the geographical criteria described above, and of the decision criterion, we obtained two output, one visual and one textual. The visual output is an ordinal decision map, based on decision rules, which allowed us to identify priority areas of intervention. In particular, we were interested in agro-environmental measures for hydro geological protection. The text output reports all the rules on which the map is based. Through two additional modules, *r.to.drsa e r.in.drsa*, it was also possible to integrate the analysis with the traditional DRSA software, as 4emka2, jMAF or Jamm, to control the elaboration by means of other algorithms not present in our module.

- [1] BLASZCZYNSKI, J., GRECO, S., MATARAZZO, B., SLOWINSKI, R., AND SZELAG, M. 2009. jMAF- Dominance-based Rough Set Data Analysis Framework- <http://www.cs.put.poznan.pl/jblaszczyński/Site/jRS.html>
- [2] CHAKHAR, S., AND MARTEL, J.M. Enhancing Geographical Information Systems Capabilities with multi-criteria evaluation functions. *Journal of Geographical Information and Decision analysis* 7 (2003), 47-71.
- [3] DENSHAM, P.J. Spatial decision support systems. In D.J., MAGUIRE, M.F., GOODCHILD AND D., RHIND (eds), *Geographical information systems: principles and applications* (London, UK, 1991). Longman, vol. 1: 403-412.
- [4] ESTALRICH, J., AND TRILL, J. GATAGRASS: A Graphical user interface for using GRASS GIS. *Computers & Geosciences* 24 (1998), 501-506.
- [5] FRIGERI, A., HARE, T., NETELER, M., CORADINI, A., FEDERICO, C., AND OROSEI, R. A working environment for digital planetary data processing and mapping using ISIS and GRASS GIS. *Planetary and Space Science* 59 (2011), 1265-1272.
- [6] GRECO, S., MATARAZZO, B., AND SLOWINSKI, R. Rough approximation of preference relation by dominance relations. *European Journal of operational research* 117 (1999), 68-83.
- [7] GRECO, S., MATARAZZO, B., SLOWISKI, S., AND STEFANOWSKI, J. An algorithm for induction of decision rules consistent with the dominance principle. In ZIARKO, W. AND YAO, Y. (eds), RSCTC, (2000, Heidelberg, NL). Springer-Verlag, pp 304-313.
- [8] GRECO, S., MATARAZZO, B., AND SLOWINSKI, R. Rough sets theory for multicriteria decision analysis. *European Journal of operational research* 129 (2001), 1-47.
- [9] HINLOOPEN, E., NIJKAMP, P., AND RIETVELD, P. Qualitative discrete multiple criteria choice models in regional planning. *Regional Science and Urban Economics* 13 (1986), 77-102.
- [10] JANKOWSKI, P. Integrating geographical information systems and multiple criteria decision- making methods, *International Journal of Geographical Information Systems* 9 (1995), 251-273.
- [11] LAARIBI, A., CHEVALLIER, J.J., AND MARTEL, J.M. A spatial decision aid: a multicritererion evaluation approach. *Computers, environment and urban systems* 20 (1996), 351-366.

- [12] LASKAR, A. *Integrating GIS and Multi Criteria Decision Making techniques for land resource planning*. International Institute for Aerospace survey and earth sciences (ITC), Enschede, 2003.
- [13] LI, X., DI, L., HAN, W., ZHAO, P. AND DADI, U. Sharing geoscience algorithms in a Web service- oriented environment (GRASS GIS example). *Computers & Geosciences* 36 (2010), 1060-1068.
- [14] MALCZEWSKI, J. *Gis and Multicriteria Decision Analysis*. Wiley, New York, 1999.
- [15] MALCZEWSKI, J. Gis- based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science* 20 (2006), 703-726.
- [16] NETELER, M., AND MITASOVA, H. *Open Source GIS- A GRASS GIS approach*, third ed. Springer, New York, 2008.
- [17] NIJKAMP, P., AND HINLOOPEN, E. Qualitative multiple criteria choice analysis, the dominant regime method. *Quality & Quantity* 24 (1990), 37-56.
- [18] ROY, B. The outranking approach and foundations of ELECTRE methods. *Theory and decision* 31 (1991), 49-73.
- [19] ROY, B. *Multicriteria methodology for decision aiding*. Kluwer Academic Publishers, Dordrecht, 1997.
- [20] SAATY, T.L. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 15 (1977), 234-281.
- [21] YAGER, R.R. Multiple objective decision-making using fuzzy sets. *International Journal of Man-Machine Studies* 9 (1977), 375-382.
- [22] YAGER, R.R. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics* 18 (1988), 183-190.
- [23] YAGER, R.R. Families of OWA operators. *Fuzzy Sets and Systems* 59 (1993). 125-148.

How to make R, PostGIS and QGIS cooperate for statistical modelling duties: a case study on hedonic regressions

Author

Olivier Bonin, Université Paris Est - IFSTTAR - LVMT, France

KEYWORDS : return on experience, statistical modelling, visualisation, R, PostGIS, QGIS

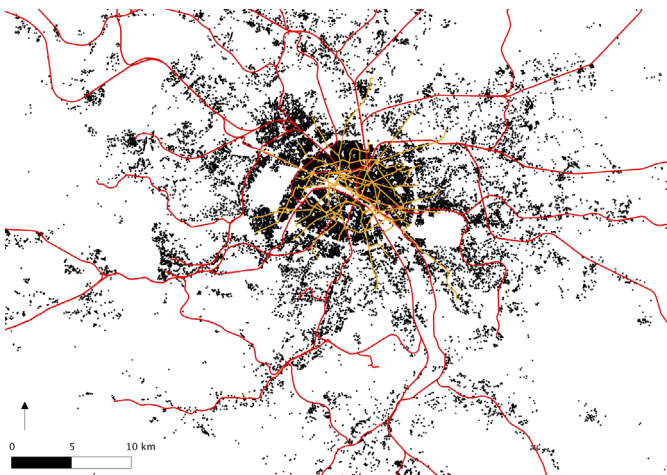
Hedonic regressions used to model housing prices are a good example of statistical modelling that is demanding towards software: housing databases are generally large (typically several hundred thousands of records), and many spatial characteristics must be computed for each housing location. An hedonic model for housing prices [3, 8] is generally written as :

$$p_i = \sum \alpha_j x_{ij} + \sum \beta_j y_{ij} + \sum \gamma_j z_{ij} + e_i$$

where x_{ij} denotes structural characteristics of housing i (e.g. size, number of rooms, presence of an elevator, etc.), y_{ij} the neighbourhood characteristics (e.g. presence of rapid transits in the vicinity), z_{ij} the market position characteristics (e.g. distance to employment, distance to the city centre), p_i the housing prices (or more generally Box-Cox transforms of these prices), and e_i a Gaussian error term.

The structural characteristics are found in housing price databases. The location of each housing consists in a postal addresses or in geographical coordinates. The spatial characteristics and are computed with the help of spatial queries as well as with network analyses. Thus it is necessary to be able to load road networks and public transportation networks and perform classical network analyses such as shortest path computation. It is also necessary to manage several layers of spatial information such as the location of employment centres and of amenities.

To select pieces of software to perform our modelling duties, we have to keep in mind that we want to navigate within two distinct worlds: the world of statistics and the world of geomatics. On one hand, a statistical software is mandatory to perform model estimation. R [5] is probably the best candidate for this task, even compared to commercial systems, given that our modelling might include spatial regression techniques or multi-level modelling [1]. On the other hand, a geographical information system (or a spatially-aware database system) is useful to perform spatial analysis (though R can perform part of these tasks thanks to some of its libraries). Here, the number of mature open source geographical information systems is relatively high, and several projects have reached the state to be able to cope with the hundred thousands of points associated to the location of dwellings. And globally, as a large amount of data is involved, it may prove useful to store and access it through a database management system. PostGIS [6] is a very good and easy to use spatially-aware RDBMS. Moreover, the spatial index in PostGIS will be helpful to perform spatial queries on hundred thousands of points. While R and PostGIS are obvious choices, the choice of the candidate geographical information system is very open. QGIS [4] has been selected in the present study because it has the ability to connect to both PostGIS and R natively or with the help of readily available plugins, and has decent mapping capabilities (Figure1).

**FIGURE 1**

Rapid transit network and housing locations in the Ile-de-France region (map drawn with QGIS, data from the BIEN database)

We wanted our system to be able to run on Windows, GNU/Linux and Mac OS X, which is the case for the three softwares. However, in our setup with a PostGIS server on GNU/Linux and R and QGIS on Mac OS X, we have discovered that some of the required libraries required to make software communicate could be a bit tricky to install and configure on Mac OS X.

We describe in this communication how we managed to make R, PostGIS and QGIS work together to assist us in our econometric modelling of housing prices. We have evaluated different combinations of these three pieces of software that we have used extensively during the course of this research, to finally converge to the system that we are currently using. We were concerned mainly with performance (as large amounts of data are involved), spatial query capabilities, easiness to compute new spatial indicators and to take them into account in the statistical modelling, and finally spatial visualisation of statistical results and map production. As it is often the case with open source software, several possibilities exist to connect the pieces of software two by two. We tried some of them and give an insight on the results of our experiments.

We chose PostGIS as a central repository for all our tabular and spatial data, and R as statistical system. To our knowledge, R can be connected to PostGIS either by “RODBC” [7] or by “Rdbi” and “RdbiPgSQL”. Rdbi (hosted on BioConductor [2] and thus accessible through this repository) proved to be much faster than RODBC in our case, so that we have settled on it. Moreover, having ODBC work on Mac OS X requires to compile the PostGIS driver and to configure the connexion by hand. To connect QGIS to PostGIS was performed with QGIS native driver. We have also connected QGIS directly to R for testing purposes. This connexion can be established by the “manageR” plugin (and probably other plugins). It did work, but did not prove to be very useful: because of the size of the housing price database, loading data from files in QGIS exhibited horrible performance compared to PostGIS. The performance of the final system is quite good. Data can be exchanged between R and PostGIS in both directions with decent times, and from PostGIS to QGIS for visualisation purposes. We almost never had to exchange data from QGIS to PostGIS, as all spatial queries were directly performed in PostGIS by SQL queries.

Our return on experience on these three pieces of open source software deals with statistical analysis, and more precisely the modelling of the influence of accessibility on housing prices. These kinds of models can require somewhat sophisticated statistical methods when spatial auto-correlation of residuals must be controlled and taken care of. Of utmost importance to us was the possibility to visualise our results on maps, so that we focus here on the visualisation duties.

The typical modelling process is to import data from PostGIS into R, estimate models, add the model error term into the data frame used for modelling and then update the PostGIS database with the new values. It is clear from Figure 1 that mapping the error term at the point level leads to hard to read maps, because of point density and because several dwellings share the same coordinates as the result of the (documented) geocoding process of dwellings in the BIEN database. Thus we choose to work on aggregate error terms, at the IRIS level or at the town level. Figure 2 is an example of such a map at the IRIS level on Ile-de-France.

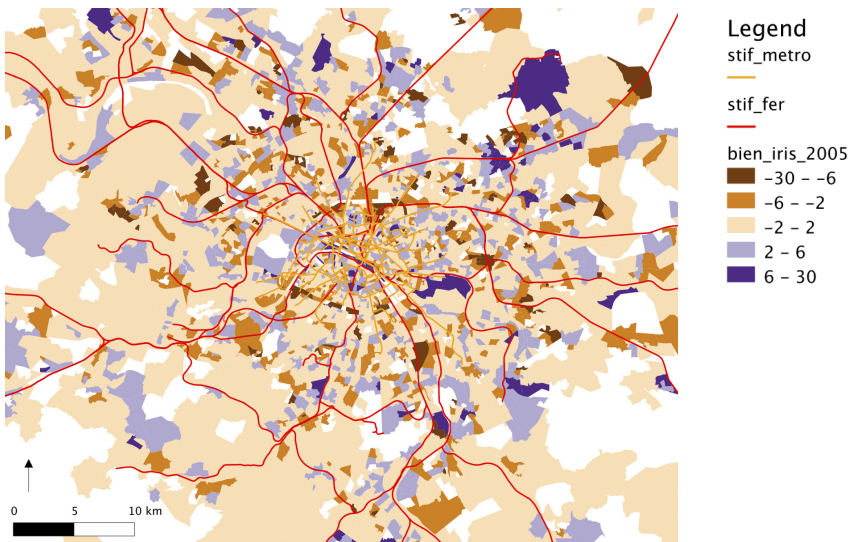
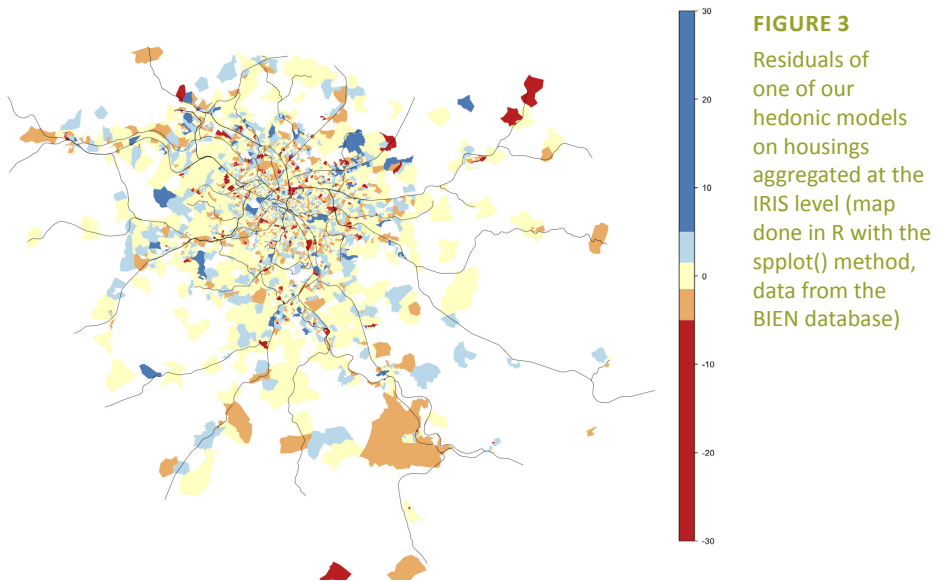


FIGURE 2

Residuals of one of our hedonic models on housings aggregated at the IRIS level in Ile-de-France (map done with QGIS, data from the BIEN database)

However, with our progressive discovering of R spatial handling facilities, we use more and more R and PostGIS without QGIS. Indeed, we are able to load shapefiles, perform analyses and thematic maps with the help of the “sp” and the “maptools” libraries. The shapefiles describing public transit networks, employment centres or administrative boundaries are small enough to be efficiently handled by R. Thus, we obtain quicker maps of error terms, probably cruder than the ones we can elaborate with QGIS, but sufficient for analysis and even for publication.



The most striking conclusion of our experiment is that, despite the strong requirements of our research in spatial analysis, the geographical information system appeared to be the most dispensable piece of software. QGIS showed irritating limitations when working on PostGIS tables (by comparison to working on shapefiles), and its visualisation module worked decently but showed real ergonomics problem when tuning cartographic representation. Of course there might have been better choices than QGIS, but our experiments

with other geographical information systems were not better.

We were able to progressively put aside the geographical information system in our setup because of the excellent spatial processing capabilities of PostGIS, and to the development of a large number of spatial functions inside R, including the ability to load shapefiles and compute spatial queries. Our conclusion on working on the interface between statistics and geomatics is that statistical systems such as R are bridging the gap between both worlds. In our opinion, geographical information systems have to work a great deal towards statistical analysis to regain the interest of researchers performing quantitative analyses in social sciences.

[1] BONIN, O. Urban location and multi-level hedonic models for the Ile-de-France region, *ISA International Conference on Housing*, Glasgow, 2009

[2] BIOCONDUCTOR. Open source software for bioinformatics, retrieved September 6, 2012 from <http://bioconductor.org/>

[3] KAIN, J. F., AND QUIGLEY, J. M. Measuring the value of house quality. *Journal of the American Statistical Association* 65(330): 532-548, 1970

[4] QUANTUM GIS PROJECT. Welcome to the Quantum GIS Project, retrieved September 6, 2012 from <http://www.qgis.org>

[5] R Development Core Team. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2009

[6] REFRACTION RESEARCH. What is Postgis, Postgis : Home, retrieved September 6, 2012 from <http://postgis.refrations.net/>

[7] RIPLEY B., AND LAPSLEY, M. RODBC: ODBC Database Access, CRAN - Package RODBC, retrieved September 6, 2012 from <http://cran.r-project.org/web/packages/RODBC/index.html>

[8] ROSEN, S. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82(1): 34-55, 1974

Toward a coupling between GIS and agent simulation. USM: an OrbisGIS extension to model urban evolution at a large scale

Authors

- **Frederic Rousseaux**, UMR LIENSS - ULR/CNRS, *France*
- **Erwan Bocher**, IRSTV, *France*
- **Antoine Gourlay**, IRSTV, *France*
- **Gwendall Petit**, IRSTV, *France*

KEYWORDS : Urban environment, simulation, agent-based modeling, spatial analysis, GIS, open-source

Introduction and context

The modeling of a city's evolution is one of the most exciting fields in urban studies research. The interest of such a model lies more in the past than the future: understanding how a city was made rather as opposed to how it might evolve. Such a model can necessitate expertise from numerous fields. From urban network analysis to spatial organisation and social patterning, modeling is a matter of choice. Our model simulates a set of alternative planning scenarios and urban development options. City growth reveals important issues to be taken into account in the city planning process and in territorial diagnoses, and analysis of potential alternatives is an essential step for both technicians and policy makers.

Using a spatio-dynamic urban evolution model, we discuss the implementation of agent-based modeling in an open-source geographical information system (GIS). Our model simulates the evolution of urban structures as a function of different scenarios at the city block scale in Nantes-Métropole, France.

From an agent-based paradigm to a GIS representation

A spatial simulation necessarily describes a given phenomenon and the space in which it is embedded. This description requires appropriate (and often very specialized) software. One of the main technical challenges is to link geographical information systems with simulation platforms. GIS, with the richness of related tools—geographical databases, spatial analysis, cartographic representations—are needed to realistically describe the spatial environment of the simulation, but do not possess dynamic capabilities. On the other hand, agent-based modeling platforms allow simulation of phenomena in a dynamic context but lack geographic and spatial analysis tools. Agent-based models have an individual-centric structure that identifies the active entities (the agents) and defines their behaviour, which can be influenced by interactions with other agents as well as the environment. *“The environmental model is critical for the Multi-Agent System as it strongly affects the agents’ decision making and behaviour.”* [1]

Coupling spatial capabilities (e.g., managing georeferenced vector data) and agent based simulation platforms (e.g., assigning behaviour and rules to geographic features) has already been implemented in open-source toolkits such as REPAST S, based on the geotools library [2], GAMA [3], or the simpler NETLOGO [4] and its GIS extension. In this paper, we propose the description of an agent-based model implementation in the open-source GIS platform OrbisGIS [5]. This platform is characterised by high-performance processing, easy handling on important geographical databases as well as complete cartographic representation capabilities. OrbisGIS stores geographical data in a GDMS database and performs spatial analysis chains using its own language [5].

Description of the Urban Simulation Model (USM)

First, we describe in our model a geographic environment (Nantes-Métropole) at the city block scale by means of a large-scale database (BD TOPO, IGN). The model includes road networks, amenities, land-use characterizations, planning rules, and cadastral plots.

In our model we characterize two classes of agents which have their own behaviour and are capable of evolving in both time and space. Using geographic vector data in an agent-based model allows us to consider a geographical object as agent [3].

The first type of agent is linked to the cadastral plot, and is characterized as an evolving homogeneous urban structure. Each plot agent is identified with a method based on a remote sensing classification. Five categories of urban structures are defined based on density of buildings, ranging from very low to very high (Figure 1). The plot is determined according to specific evolutionary rules, based on planning rules and attractiveness. As an example, a given zone can evolve from a non-building structure to a building structure or from an existing building structure to a denser building structure depending on local urban rules and the number of households it contains.

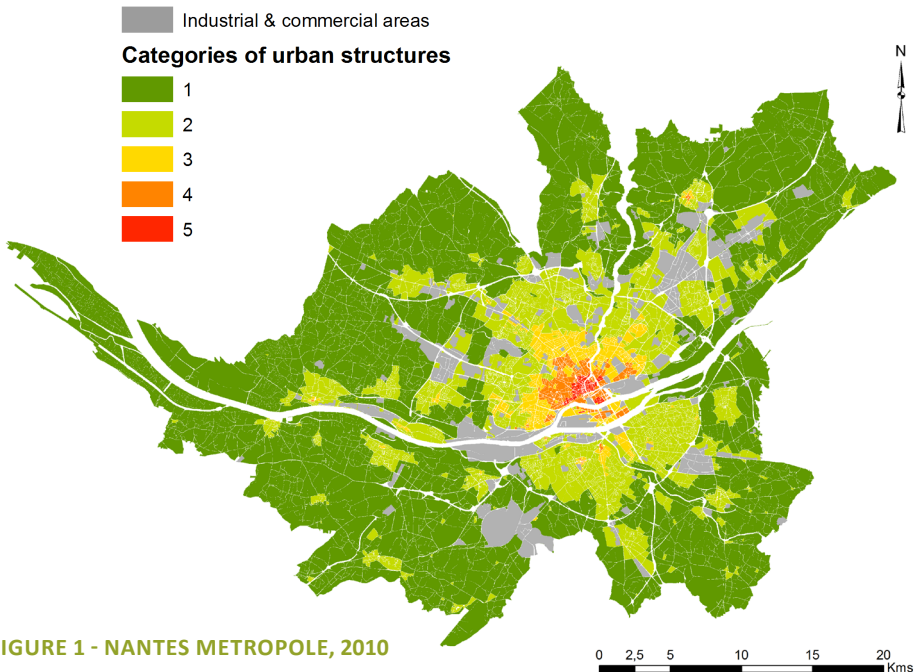


FIGURE 1 - NANTES METROPOLE, 2010

Initial step of the simulation

Households are the second type of agent. This agent is described by average age and income, factors which influence the need for space and the corresponding type of urban structure (e.g., house with a garden or building). Each household searches for a satisfactory plot to settle, considering the housing category and neighbourhood corresponding to similar incomes and the social segregation model [6]. If a given plot is not suitable, the household moves on and tries another plot. The more attractive and easier build a plot is, the faster the urban structure becomes dense.

The massive number of agents in the model (100 000 households, 60 000 plots) and the high level of geographical environment detail require a very high-performance platform with complete GIS capabilities. We developed our model from scratch as a plugin in OrbisGIS.

USM, an OrbisGIS plugin

OrbisGIS is an open-source GIS written in Java and developed at the IRSTV (Institut de Recherche Sciences et Techniques de la Ville) in Nantes, France since 2007 [5]. The power of OrbisGIS lies in its GDMS (Generic DataSource Management System) library, which offers its own SQL language and an abstract data layer to read, write and process huge amounts of data [7]. GDMS implements the Simple Feature SQL standard to perform data processing with spatial operators (ST_Buffer, ST_Intersects, etc.). All spatial analysis methods are expressed with SQL functions, just as in PostGIS#. Moreover, GDMS offers an application programming interface (API), and OrbisGIS is a graphical interface that allows creating custom processes and user interfaces. The USM model takes advantage of these two aspects. It is structured in two main components:

- The first part is an object-oriented implementation of the model described above, using GDMS files to store results, to read input configuration files and data as well as to run all geographic processes. Its API can be used directly to run simulations and also allows users to provide custom algorithms for some parts of the computation.

FIGURE 2
Frame to setup the USM
configuration parameters

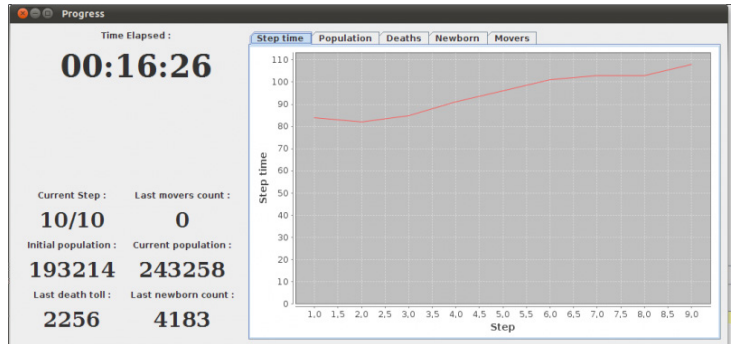
Parameter	Value
Year	2000
Number of Turns	1
Buffer size	10
Amenities Weighting	1
Constructibility Weighting	1
Idealhousing Weighting	1
Gauss Deviation	0.1
Model Selection	Statistical Model (selected)
Household Memory	5
Moving Threshold	20
Segregation Threshold	0.8
Segregation Tolerance	0.3
Immigrant Number	5000
Threshold 1	0.000000
Threshold 2	0.000155
Threshold 3	0.001000
Threshold 3	0.001466

Please select a target file

→ The second part is a graphical interface accessible through OrbisGIS using the Java Swing GUI library. The USM user interface allows one to configure the model or open a previously configured one, and then to run it (Figure 2).

During the simulation execution, the state of several key parameters is visible using dynamic charts, including the number of incoming, outgoing and total households (Figure 3). Each step is stored in a separate GDMS file and can be linked to the input parcel data to display the result of the simulation as a thematic map. For users, a complete visual wizard interface has been developed. All the parameters of the model's configuration can be saved. After the simulation, the user can display each step of the process in OrbisGIS, access all GIS primitives and use them for spatial analysis or dynamic cartography.

FIGURE 3
Dynamics charts
are showing
each step of the
simulation



Results and further work

Our plugin's processing time is very short considering the large number of agents. Most of the processing time is linked to the movement of agents. The model can treat very large sets of geographic data (more than 100 000 parcels) without significant increase in processing time. Nevertheless, the number of movers can increase it, albeit lightly. As an example, 5 000 agents takes 10 minutes of computation.

Concerning the case study, we tested two scenarios in Nantes-Métropole, a major metropolitan area, located near the west coast of France. Both scenarios propose a time scale of 2010 to 2030. Different categories of urban structures are numbered 1 to 5 in Figure 4. Each category is an indicator determined from a set of variables (density, building height, percent empty space, etc.). Commercial and industrial areas are a specific category that we ignore in this model due to very specific transition rules.



FIGURE 4
categories of urban structures

FIGURE 5 - NANTES METROPOLE, 2030

Results of scenario 1

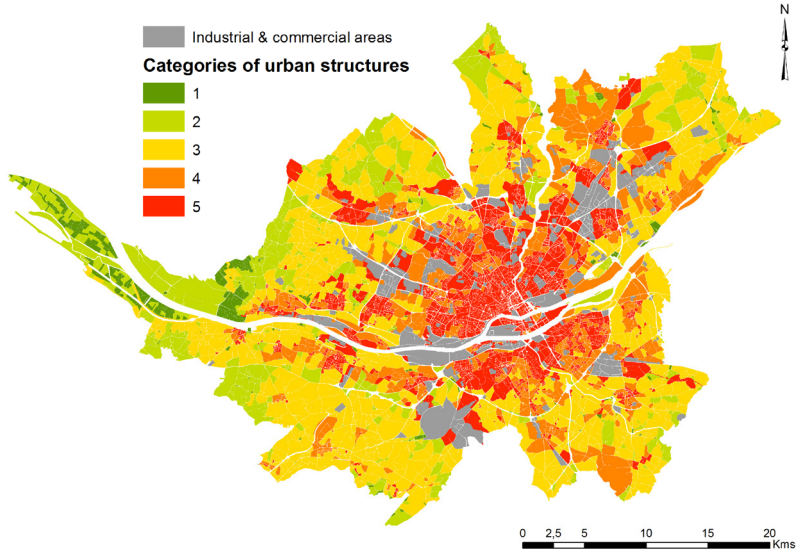
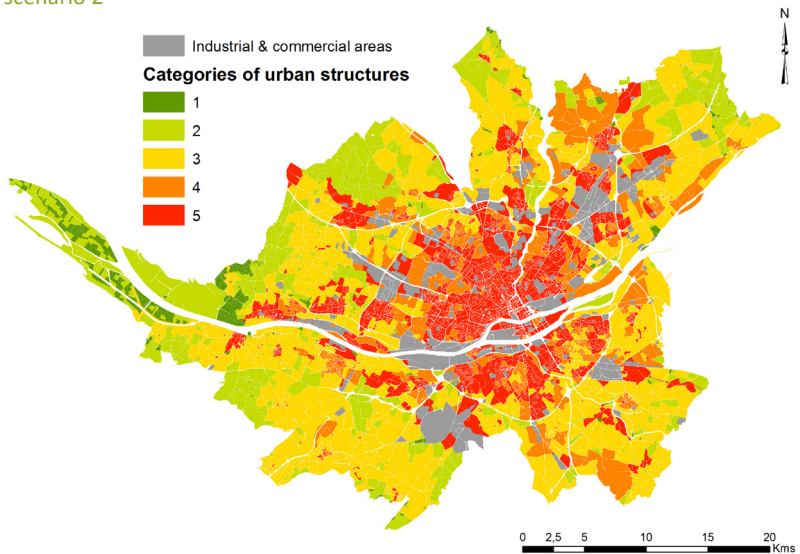


FIGURE 6 - NANTES METROPOLE, 2030

Results of scenario 2



The first scenario (Figure 5) simulates a political desire to curb urban sprawl and favours densification. The second scenario (Figure 6) does not constrain urbanization and “let’s the city live”. As a consequence, in 2030 the model shows important urban sprawl, to the detriment of agricultural and natural areas, especially on the northwest area of the city.

Comparing the two maps reveals spatial disparities, particularly in the southwest area of the metropolis where the structure of the plot varies from high densities (mainly orange) to low densities (mainly yellow). A similar trend can be observed in the northwest where structures with lower densities appear. By contrast in the vicinity of the city center (located in the center of the map), structure plots are denser in scenario 1 as compared to scenario 2.

One of the major interests of this GIS model is to view the results of a simulation in light of demographic data or other spatial information (e.g., land uses, traffic network distribution). The relevance of the model can then be evaluated in relation to the territory and how it changes. This is the most interesting aspect of the model; the results of the simulation can be discussed and improved upon with city planners and urban engineers.

Acknowledgments

This research is granted by VegDUD project, funded by the French Research Agency (Agence Nationale de la Recherche, ANR) under contract ANR-09-VILL-0007. The research partners of VegDUD are:IRSTV, IFSTTAR, Plante et Cité, LaSie (Univ. of La Rochelle), LPGN (Univ. of Nantes), CNRM (Météo France), ONERA, IRSN, CSTB, EPHYSE (INRA).

[1] SCHULE, M., HERRIER, R. & KLUG, F. 2004 Coupling GIS and multi-agent simulation towards infrastructure for realistic simulation. *Proceedings of MATES 2004*, vol. 3187, (Berlin, Germany, 2004) Springer, pp. 228–242.

- [2] NORTH, J., AND MACAL, C. M. Managing Business Complexity. Discovering Strategic Solutions with Agent-Based Modeling and Simulation. 2007.
- [3] TAILLANDIER, P., DROGOUL, A., VO, D.A. AND AMOUROUX, E. GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. *Proceedings of The 13th International Conference on Principles and Practices in Multi-Agent Systems*, (India, 2012), Volume 7057/2012, pp. 242-258.
- [4] WILENSKY, U. 1999. NetLogo. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, USA*. Retrieved August 28, 2012 from <http://ccl.northwestern.edu/netlogo/>
- [5] BOCHER, E., PETIT, G. «*OrbisGIS : Geographical Information System designed by and for research*», in *Innovative Software Development in GIS*, ISTE & Wiley, «Geographical Information System» series, Bucher B. & Le Ber F. (Editors), (p25-66), 331p. (2012).
- [6] SCHELLING, T C. *Dynamic Models of Segregation*. *Journal of Mathematical Sociology* (1971) Vol1, pp. 143-186.
- [7] BOCHER, E., NETELER, M. *Geospatial Free and Open Source Software in the 21st Century*, *Lectures Notes in Geoinformation and Cartography (LNG\&C) series*, Springer-Verlag.(2009)

Using open-source tools for the simulation of urban transportation systems

Authors

- **Quoc Tuan Nguyen**, L3I laboratory - university of La Rochelle, *France*
- **Alain Bouju**, L3I laboratory - university of La Rochelle, *France*
- **Pascal Estraillier**, L3I laboratory - university of La Rochelle, *France*

KEYWORDS : Intelligent transportation systems, Multi-agent, Regulation, Adaptive, Software architecture, Open-source software, Geolocation

Urban transportation systems are defined by means of urban transport plans and territorial coherence schemes. Organizational and regulatory policies concerning urban transportation systems are governed by general principles corresponding to strategic direction and policies and integrate infrastructure constraints. Transportation systems are increasingly complex and must evolve to incorporate components of sustainable development. It has become appropriate to develop high-level simulation tools for urban transportation policy makers so that they can analyze the potential consequences of their choices.

Nowadays, the open-source tools are used increasingly in many fields especially in the education, the research and applied domains. They have many advantages such as the availability of the source code and the right to modify it, the right to redistribute modifications and improvements to the code, the right to use the software in any way, etc.

In this paper, we propose a decision-maker simulator intended to define and tune urban transportation policy (travel, parking and transportation

strategies) using exist open- source tools such as LibreOffice, OrbisGIS and GAMA platform.

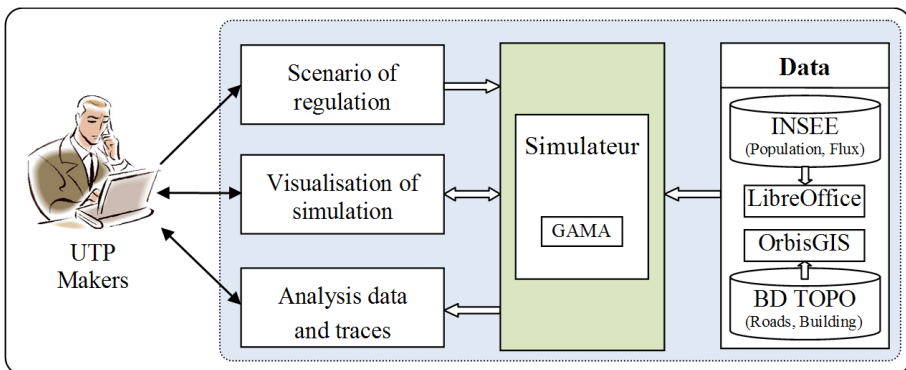
Currently, there are simulators that deal with concrete problems in the urban transportation system such as the simulation of movement of individuals [1, 2, 3, 4], driver's behavior [5], the transportation flow [6] , etc. However, there are very few simulators that address the problems of the organization of transportation systems and in particular the problem of analyzing the impacts of regulatory strategies for the transportation system.

The main objective is to provide a simulation tool to help urban transportation policy (UTP) makers to analyze and evaluate the impacts of regulatory strategies. The simulator consists of modules that contain information describing the infrastructure of transportation systems, means of transport, signaling, individuals' behavior etc. The Figure 1 illustrates the interaction between the urban transportation policy maker and simulator components and presents an overview of our system. The simulator must then take into account the following features :

- Support scenarios of regulation: this feature allows the decision maker to define scenarios of regulation strategies (set indicators for the strategies of regulation).

FIGURE 1

Overview of the system (UTP – Urban Transportation Policy)



- Simulate movement of individuals in the context of multi-modality of transport: this feature simulates multimodal travel corresponding to a plan of activities (work, school, leisure, shopping, services, etc) of individuals on the infrastructure of the transportation system. The simulation must be multi-scale time and space.

The problems of our research contribute to:

- the development of transportation systems: a theoretical framework for modeling, software architecture supporting configuration and adaptive regulation and integration of temporal and spatial aspects.
- the development of a decision support simulator: modeling of transportation infrastructure, simulation of user's behavior and analyzing the impacts of regulatory strategies.

In terms of performance, simulators organizing transportation systems based on the multi-agent systems paradigm are well suited to complex dynamic systems and can describe the behavior of real systems for which equation models are not always satisfactory, particularly when an algorithmic approach is preferred to a probabilistic approach.

In terms of system architecture, we adopted a “system of systems” approach [7, 8], mainly structured in layers, in order to the main elements of the system. In our proposal, each layer plays a role as a system. We represent explicitly, for example, a layer of roads, lights, parking, means of transport, etc. Our system uses an agent-based simulation incorporating spatial and temporal information. It must support the regulatory scenarios to simulate the effect of regulatory strategies on transportation systems.

The input data of the simulator are geographic data for the study area, infrastructure of the transportation network (land, roads) provided by the BD TOPO 2 from IGN (the French National Geographical Institute), the results of surveys and the general census of the population provided by INSEE that contain the information to set individual's behaviors. The advantages of this INSEE (the French National Institute of Statistics and Economic Studies) source consist in the fact that this is a database with large sample, which ensures accurate data and provides a spatial presentation of the population, and data movements of individuals.

In addition, we implemented the mechanism of «traces»; the trace files contain the result of simulation. Travel surveys, census and traffic measurements were used. Analysis of available data and traces were used to evaluate the suitability of our simulations according to different regulatory strategies.

Finally, we implemented a prototype for the movement of people in the city of La Rochelle – France with data from INSEE 2006 and BD TOPO 2 (delivered on 22/10/2011).

LibreOffice is the power-packed free, libre and open source personal productivity. We used Calc and Base tools of LibreOffice to analyze and manipulate the data of INSEE in order to categorize individual's profile and the mobility flux.

OrbisGIS is a Geographical Information System dedicate to scientific modeling and simulation. OrbisGIS is developed by the Institute on Urban Sciences and Technics – France (IRSTV - CNRS/FR-2488), it is an open-source software. We used the OrbisGIS to process the geographical data of BD TOPO 2 for system infrastructure (roads, paths), attraction points (home, work, school, etc.).

GAMA [9, 10] is a simulation platform which aims at providing field experts, modelers, and computer scientists with a complete modeling and simulation development environment for building spatially explicit agent-based simulations. It is developed with the Java programming language and is open source. Geographic information is well integrated. It has implanted example models and has a language for defining models.

We installed a simulation of individual movements using the GAMA simulation platform version 1.4. System infrastructures (roads, paths, buildings) were stored in shapefiles (Esri ref). Each section of road or path was represented by an agent, each building was also represented by an agent and all these manipulations were performed automatically from a shapefile. The number of individuals in this simulation was 10000 (the actual population of La Rochelle is around 76000). Each individual was represented by an agent, it had a place of residence and a place of work or study, its activity plan consisted only of the round trip from home to work or study. The individual movements were determined by the shortest path algorithms. The results

obtained proved the feasibility of our choices for the design of our simulator and for the integration of geographic information (roads, buildings) in the simulation.

-
- [1] LANG, C., BANOS, A. AND THEVENIN, T. *Représentation dynamique et visualisation interactive de données spatio-temporelles dans le projet MIRO*. Projet MIRO, 2006.
- [2] LANG, C. AND al. MIRO, *Rapport Final Septembre*, 2007.
- [3] KOOHBANANI, M. J. *Enhancements to Transportation Analysis and Simulation System (TRANSIMS)*. PhD thesis, 2004.
- [4] NAGEL, K. Multi-modal traffic in TRANSIMS. *Traffic* (2001).
- [5] ESPIE, S. ARCHISIM : Collaboration MSIS / LPC modélisation et simulation du comportement des conducteurs. MSIS, 2004.
- [6] BEHRISCH, M. et al. 2011. SUMO – Simulation of Urban MObility. SIMUL 2011: *The Third International Conference on Advances in System Simulation* (Barcelona, Spain, 2011), pp. 63-68.
- [7] LUZEAUX, D., & RUAULT, J.-R. *Ingénierie des systèmes de systèmes - concepts et illustrations pratiques*. Hermes Science. 2008.
- [8] MAIER, M.W. Architecting Principles for Systems of Systems. *Systems Engineering 1* (1998), pp. 267-284.
- [9] DROGOUL, A. *An introduction to GAMA Reasons for developing GAMA Original purpose of GAMA*. Training session on agent-based modeling, Hanoi, VietNam, 2009.
- [10] TAILLANDIER, P., DROGOUL, A., VO, D.A. AND AMOUROUX, E. GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In *'The 13th International Conference on Principles and Practices in Multi-Agent Systems (PRIMA)'* (India, Volume 7057, 2012), pp. 242-258.

GeoPeople project: using RESTful Web API to disseminate geohistorical database as open data

Authors

- **Eric Grosso**, IGN - French Mapping Agency, *France*
- **Christine Plumejeaud**, IGN - French Mapping Agency, *France*
- **Benjamin Parent**, IGN - French Mapping Agency, *France*

KEYWORDS : GeoPeople, demographic evolution, Web API, geohistorical database, old maps, France

Introduction

For 250 years the French territory has considerably densified throughout population growth and technological advances, thus changing the topography of the French landscape. Moreover, industrialisation has generated a significant rural exodus. The study of the relationship between the evolution of the French landscape and the French population distribution evolution is relevant, both from a historical point of view and from a land settlement point of view: if the past is better understood, it is possible to better anticipate future needs. But this topographic past has been transmitted to us in the form of old maps (for example Cassini maps from the eighteenth century and Etat-Major maps from the nineteenth century), which offer a codified trace following the map specifications relative to their eras.

The GeoPeople research project aims to analyse and extract by vectorisation the contents of old maps – Cassini maps (1:86,400 scale), Etat-Major maps (1:80,000 scale) and 1960's topographic maps (1:25,000 scale) – and to use current topographic databases in order to build a geo-historical (or spatio-temporal) database, and to study the densification of French territories. This project, funded by the ANR (French National Research Agency), brings together different partners: the COGIT laboratory of the French

Mapping Agency (IGN), the LaDéHis (historical demography laboratory) of the EHESS (School for Advanced Studies in the Social Sciences) and the MALIRE team of the LIP6 (Paris 6 laboratory in computer science).

Building such a geo-historical database requires different processes such as old maps scanning, identification of concepts and their organisation represented in an appropriate ontology, vectorisation (manual and automatic), georeferencing or data matching. We focus here on the diffusion process of the gathered data via an open Web server which allows for both an interactive exploratory spatio-temporal data analysis process of the change (administrative, demographic as well topographic) following precepts detailed in [1] or [2] and a free access to geo-historical data.

A geo-historical database gathering 200 years of evolutions

The geo-historical database produced during the GeoPeople project contains demographic data (34 census since 1793 up to 2006), historical data describing the story of administrative entities and associated topographic data covering a 250 year period (since 1750 to present). The administrative entities in question are French municipalities (“communes” in French) and their upper entities. For all of them, the database has been designed to handle their change of name, border, membership to upper unit, or administrative center, all sorts of changes that happened frequently and have led to the disappearance of 7,300 communes from a total account of 44,000 since their date of creation in 1793. Changes are modeled as events that can occur either on a single entity (change of name for instance) or simultaneously on many entities, leading to the disappearance of entities by fusion inside others or the creation of new ones by land plot transfers of pre-existing entities or the division of pre-existing entities. Furthermore the design is not limited to administrative entities: it can handle various nomenclatures (religious, academic, fiduciary, judicial, etc.) and also particular cases such as the multiple co-belonging to various entities at the same time for one entity. In order to handle topographic data (roads, rivers, buildings, fabrics, mills, churches locations) collected throughout the digitalisation of Cassini maps (1759-1789) and Etat-Major maps (around 1890) as well as the border definition changes, this new database is now also a spatio-temporal one.

This is a new development with regards, for example, to the INSEE database [3], or to previous databases, whose content was explained in [4]. The current one is focused only on historical and demographic data and is connected to a Web interface [5]. Its success has proved the great interest that French citizens have for their villages' history. However this previous Web interface does not give a free access to these data nor does it allow for a real interactive exploratory data analysis process of this history in conjunction with topographic and demographic data.

The amount of work represented by the digitalisation of maps (even though limited to 4 maps covering 80 km by 50 km, for Saint-Malo, Reims, Grenoble and Agen locations) as well as building the map of French administrative entities' history is considerable, time consuming and very valuable for scientists such as historians, demographers or ecologists. Overall it has been funded with public funds and this is why the different partners of the project considered that these data are due to be accessible to all French citizens and decided to share this information, raising thus the question of the means of this diffusion.

Towards an open data diffusion allowing easy data handling and sharing

To answer this question, a first step is to provide everyone with free downloadable data, falling within the current open data context. However, even if data provided in this way is necessary, it is not sufficient. Indeed, this solution does not generally allow users to handle this data dynamically without reprocessing it using desktop software (mainly in order to manually recreate the dynamic links originally defined in the database management system), nor is it able to handle this data directly on the Web.

Therefore, the decision was made to create a website from which it is possible for example to search for the history of a place. Because the understanding of entity evolution is complex, based on different factors (topographic, politic and demographic ones), the results page is made of several synchronised views in four interactive tabs, with a time line similar to [6], [7] or [8]. The first tab displays topographic data at different periods, the second tab contains a systemic diagram showing the administrative entity relationships over time as shown in Figure 1, the third contains a chart showing the demographic

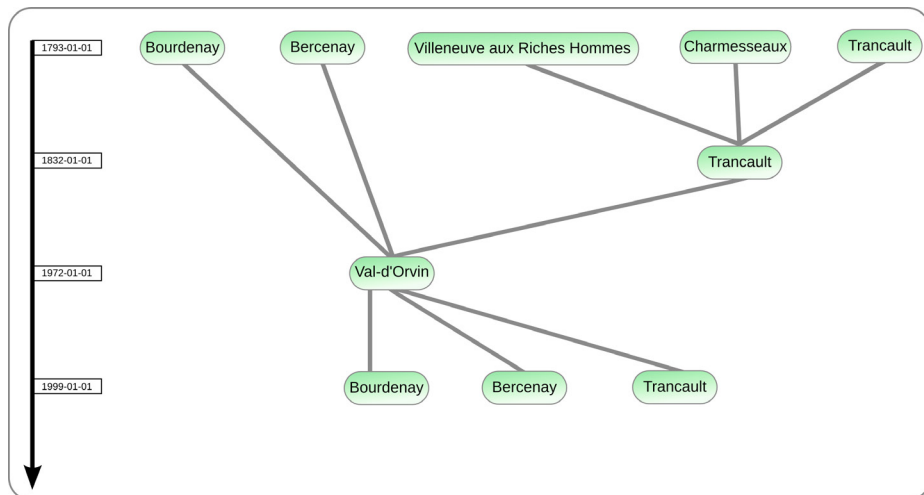


FIGURE 1

One website element : systemic diagram of Charmesseaux and its neighbouring administrative entities over time

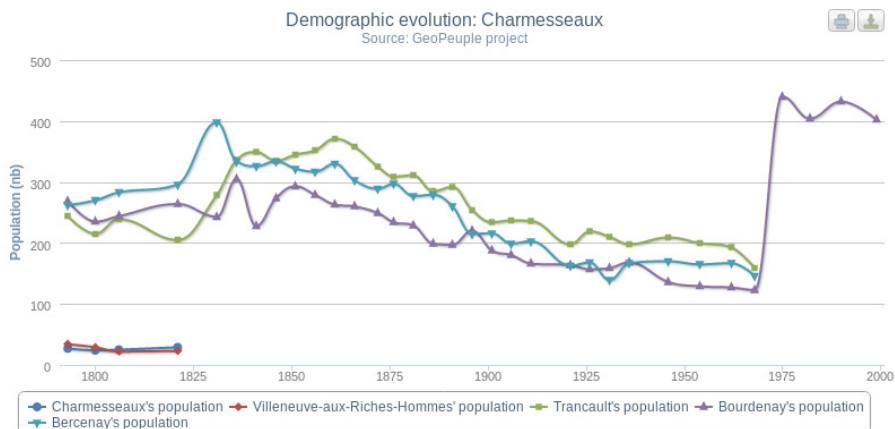


FIGURE 2

Another website element : demographic evolution of Charmesseaux and its neighbouring administrative entities over time

evolution of the studied administrative entity and its neighbouring entities as illustrated in Figure 2, and the last one combines the geometrical view of limits at various time with two other views: a demographic and systemic one presented simultaneously as proposed in [9].

This solution provides a user-friendly access to information and thus appears more flexible than just a simple download, however, it also presents two major drawbacks. The first is that a website does not allow full access to the information contained in the database, because it is limited to the original use for which the website was designed. The second is that the results displaying in HTML can not easily be handled by users. Therefore creating a mashup by crossing HTML results and external data appears quite complex for inexperienced users.

Consequently, we decided to provide a modular solution by developing a RESTful Web API which allows users to easily query the database, to serve several formats handle directly on the Web, and to build the GeoPeople

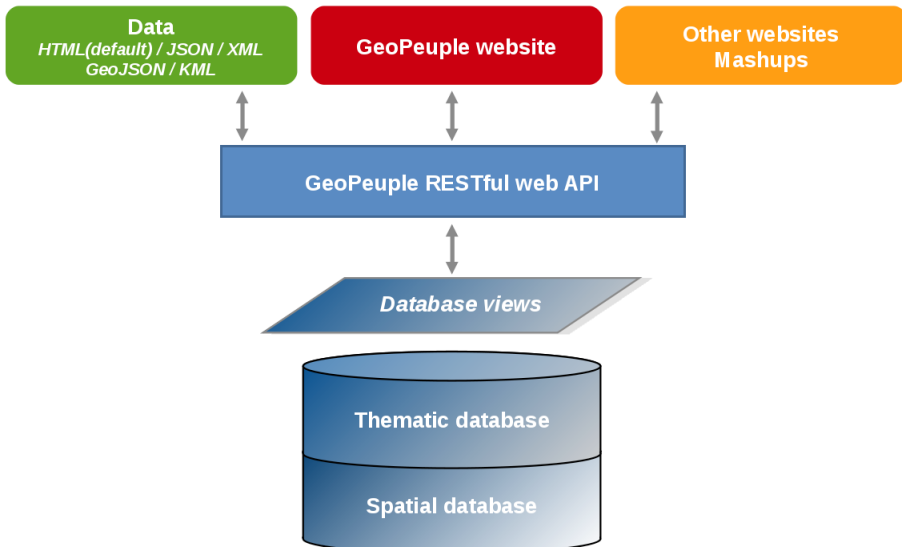


FIGURE 3

Global architecture of data diffusion

website on top. This global architecture is described in [Figure 3](#). In order to help users query the Web API or to integrate the API easily into Web pages, HTML and JavaScript (AJAX) code snippets are provided. Furthermore, the code from this Web API will be released on an open source licence at the end of the project enabling the possibility for other projects to reuse similar code architecture.

Technical API description

In order to release the code under open source licence, the full architecture including the RESTful Web API and the website is based only on open source components. On the server side, PostgreSQL and PostGIS have been chosen to store and manage the geo-historical database, Java and the Spring framework to develop the RESTful Web API, Hibernate to map the object-oriented domain model to the database, Tomcat as a Web application server and GeoServer to share spatial data through Web Map Service. On the client side, jQuery is used to develop the code snippets, Raphaël JS to create the diagram showing the administrative entity relationships over time, Highcharts to create the demographic evolution charts (Highcharts is free for non-commercial use) and OpenLayers to display maps. Finally, the RESTful Web API serves W3C and OGC standard formats such as HTML (by default), XML and KML, but also JSON and GeoJSON to ease data handling.

Conclusion

This paper presented the architecture we chose to develop in order to share the geo-historical database produced in the GeoPeople project. This architecture allows to share freely data that users can easily handle directly on the Web both by querying the RESTful Web API and by integrating HTML/jQuery code snippets directly on their own website.

A first advanced prototype has already been implemented (both server and client side). All the components used are open source allowing us also to release the RESTful Web API and GeoPeople website codes under an open source licence (GNU Affero GPL v3 license). This will be done before mid-2013. Finally and as a perspective, the RESTful Web API could easily be extended to allow users to improve the database by crowdsourcing.

- [1] ANDRIENKO, N., AND ANDRIENKO, G. *Exploratory analysis of spatial and temporal data*. Springer-Verlag, Berlin, 2005.
- [2] ANTONI, J.-P., AND KLEIN, O., AND, MOISY, S. *Cartographie interactive et multimédia : vers une aide à la réflexion géographique*. Cybergeog : European Journal of Geography (Paris, France, 2004).
- [3] Institut National de la statistique et des études économiques, 2012. History of French administrative entities. Retrieved September 7, 2012 from <http://www.insee.fr/fr/methodes/nomenclatures/cog/historique.asp>
- [4] MOTTE, C., AND SÉGUY, I., AND THÉRÉ, C. *Communes d'hier, communes d'aujourd'hui*. Les communes de la France métropolitaine. Dictionnaire d'histoire administrative. Institut National d'Études Démographiques, Paris, 2003.
- [5] MOTTE, C., AND VOULOIR, M.-C. EHESS Cassini website. Retrieved September 7, 2012 from http://cassini.ehess.fr/cassini/fr/html/1_navigation.php
- [6] PISON, G., AND, MATHIAN, H., AND PLUMEJEAUD, C., AND GENSEL, J. *Exploring world demography on line*. In Proceedings of the 25th International Cartographic Conference (Paris, France, 2011).
- [7] PISON, G., AND, MATHIAN, H., AND PLUMEJEAUD, C., AND GENSEL, J., 2011. Interactive maps of the world population. Retrieved September 7, 2012 from http://www.ined.fr/en/everything_about_population/interactive_maps/
- [8] BRETAGNOLLE, A., AND, MATHIAN, H., AND VAN HAMME, A. *Animating the cities: dynamic exploration of harmonized urban databases (United-States – France, 1800-2000)*. In Proceedings of the 25th International Cartographic Conference (Paris, France, 2011).
- [9] PLUMEJEAUD, C., AND MATHIAN, H., AND GENSEL, J., AND GRASLAND, C. *Spatio-temporal analysis of territorial changes from a multi-scale perspective*. International Journal of Geographical Information Systems, 2011.

An opensource tool to build urban noise maps in a GIS

Authors

- **Nicolas Fortin**, IRSTV FR CNRS 2488 Atelier SIG, *France*
- **Erwan Bocher**, IRSTV FR CNRS 2488 Atelier SIG, *France*
- **Judicaël Picaut**, IFSTTAR, *France*
- **Gwendall Petit**, IRSTV FR CNRS 2488 Atelier SIG, *France*
- **Guillaume Dutilleux**, CETE Est/LRPC de Strasbourg/6 Acoustique, *France*

KEYWORDS : opensource, GIS, noise map, urban, data, processing

Introduction

Within the framework of the Environmental Noise Directive (END) 2002/49/CE concerning the assessment and management of environmental noise, cities of more than 100,000 inhabitants are required to construct noise maps to be used in dening noise abatement action plans. For consistency, a preliminary evaluation (mainly by acoustic simulations) of the effectiveness and impacts of these action plans is mandatory.

The use of Geographic Information System (GIS) software is a tool that has become essential for such studies. GIS offers many spatial analysis functions and cartographic capabilities that are useful for understanding the impact of action plans on noise maps at different scales (i.e., buildings, administrative units, etc.). The connection between GIS and noise prediction was established more than two decades ago in a Dutch study on the impact of road traffic noise on bird reproduction. As early as 1986, simulations were carried out using Silence software on a large area in the Netherlands [3]. Unfortunately, little information is available on the acoustical aspect of the software. The

application at hand was clearly focused on the countryside and thus has few implications for urban areas. Today, Silence 3¹ combines GIS with the commercial Predictor software with a focus on highways rather than urban areas.

In this paper, we propose an alternative approach, implementing a noise prediction method within the OrbisGIS² software. The noise emission and propagation models are inspired from the French national method \NMPB 2008» [4, 6], but are simplified for two dimensional calculations in urban areas. The method, as well its implementation in the GIS software, are detailed in sections 1 and 2. We then present a case study in section 3 to evaluate several mobility plans in terms of noise impact. The particular interest of this approach is that we implement the model in the two-dimensional GIS software OrbisGIS and pay special attention to algorithm optimization in order to reduce computational times and resource consumption. Our method is able to produce noise maps for very large domains (around several millions of square meters) on a personal computer in just a few hours.

1. Sound prediction methods in a GIS: What and why

We model noise emission according to the French national method \NMPB 2008» [4, 6], based on the decomposition of line sources (i.e., traffic flow) into point sources, with some simplifications concerning the type and age of the road pavement, the stopping and starting road sections, vehicle kinematics, and the distribution of light and heavy vehicles over different time periods throughout the day. We obtain the traffic flow using a trac model developed by another partner of this research project. This model, based on the VISEM software [3], uses behavior surveys as input data and is fitted with in situ traffic flow data. The data are represented as a set of polylines (the road network) described by several attributes extracted from traffic flow simulations in VISEM. The model gives the traffic information (speed, type of vehicle, traffic lane, traffic flow, etc.) required for noise emission computations the eventual construction of noise maps using the prediction method.

We implement the prediction method in the open-source GIS software OrbisGIS. One of the most important advantages of free and open-source software (FOSS) in general is that it facilitates the reuse of existing code and

the merging of different software packages into new and less costly applications. In OrbisGIS, the user can see explicitly how noise maps are generated, unlike the so-called black box implementations of comparable commercial software packages [7]. OrbisGIS allows researchers to share their results and build a common platform to analyse sustainable urban development. It offers a unique way to understand the impact of human activities using a single tool.

Implementing a method to compute noise maps in OrbisGIS allows us to take advantage of GIS functionalities to merge different databases. For example, it is easy to link noise maps and the population distribution using spatial analysis techniques. The populations exposed to different noise levels can be located and estimated.

2. Implementation within OrbisGIS

We developed the noise map prediction method as a plug-in of OrbisGIS [9]. The noise plug-in is divided into two main parts. The first part is related to the evaluation of the sound source by considering both light and heavy vehicles as well as electric tram. The second part consists of the propagation model.

The OrbisGIS module *noisemap* uses the results of the traffic flow simulation to compute the corresponding long-term noise emission levels. These long-term emission levels can be for an entire 24-hour period (LDEN) or from 10:00 p.m. to 6:00 a.m. (LN). For each road geometry, this module outputs a spectral decomposition of the noise sources from 100 Hz to 5000 Hz.

Another OrbisGIS module is responsible for computing the noise propagation. At execution time, the module ignores the height of the buildings, so that there is no vertical direction by the tops of buildings. It takes as input data:

- an array of sound sources defined by their geometries (lines or points) and a sound level from each third octave band from 100 Hz to 5000 Hz provided by the first module
- an array of buildings (2 dimensions)
- a list of parameters, like the order of reflection and diffraction, wall absorption and the maximum propagation distance

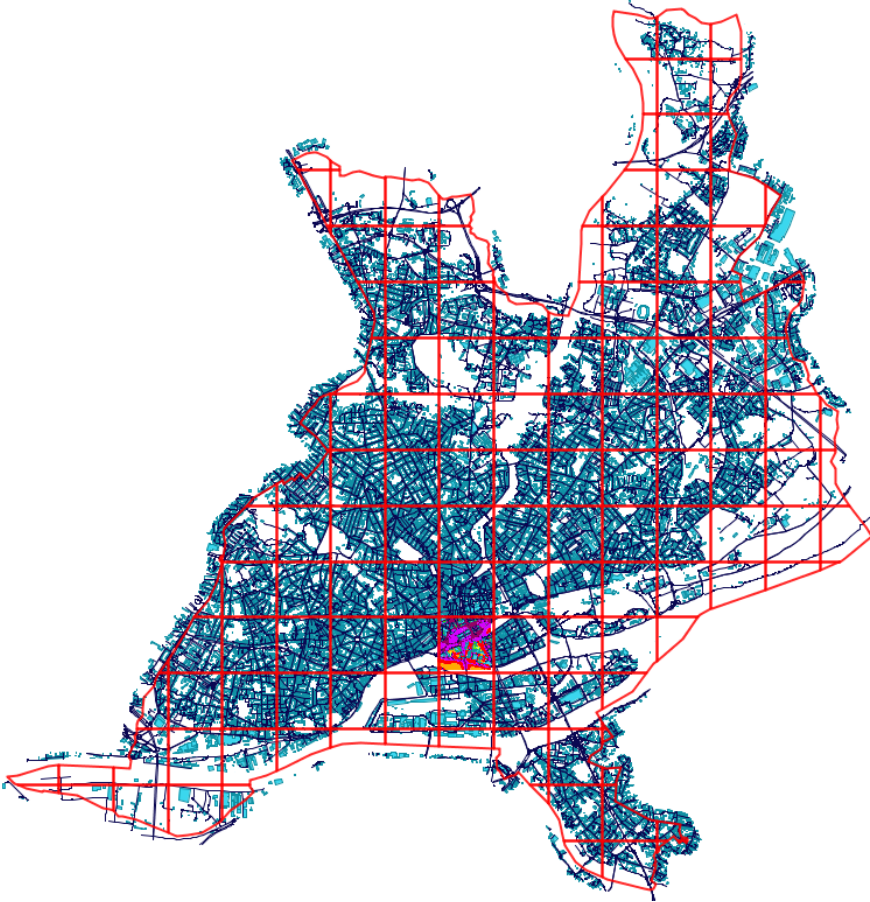


FIGURE 1

Subdomains computed according to the road network. For the colored cell :
receiver count = 253,826, direct paths = 40,256,246, reflection paths = 9,429,999,
diraction paths = 1,990,872, computation time = 3 hours

The propagation process evaluates the bounding box of the simulation area by using the full envelope of the road network. This bounding box and the maximum propagation parameter are then used to compute optimal subdivided areas of simulation (Figure 1). This leads to greater computational efficiency.

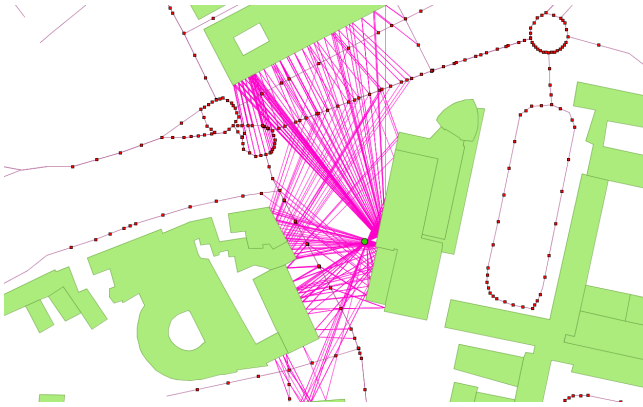
For each subarea, we form a constrained Delaunay triangulation of the union of the buildings and roads contained in the subarea (Figure 2). We place receiver points (to simulate microphones) at the vertices of the triangles. This mesh provides an optimal spatial distribution of receiver points, the density of receivers being higher near sound sources as well as buildings with a more complex geometry. An extended subarea box is defined for each subarea by taking the maximum propagation distance into account. This extended envelope does not contain any additional receivers; it is used only to enlarge the propagation domain as a function of the maximum propagation distance. In each extended subarea, outside sound sources and buildings are ignored.



FIGURE 2

Constrained Delaunay triangulation of an area of one square kilometer

At this step, the computation subarea contains all needed information: the propagation domain, the sound sources and the receivers' coordinates. We fix a particular receiver. This receiver looks for sound sources along the roads that fall within a range equal to the maximum propagation distance. Each road is then subdivided into point sources separated by a distance that depends on the distance to the receiver. When all point source locations are available, the propagation paths from the receiver to the sources are computed by considering the direct, reflected and diffracted fields (Figure 3).

**FIGURE 3**

Second reflection purple paths that connect the green receiver to the red sources

The noise contribution of each sound source is calculated, then all contributions are energetically summed into the receiver. This same calculation is repeated for each receiver in the subarea. The software then saves the noise level for each receiver in a buffer before moving on to the next subarea. Using this area subdivision approach, a region of more than 500 km² can easily be computed on a standard personal computer. In particular, we exploit multi-core technology by separating data between threads and limiting synchronisation.

It is noteworthy that the propagation module can accept a constrained list of receiver coordinates. It will then output the noise level for those receivers only. This constraint is useful for computing the noise level on building's facades. The computations are carried out by means of several SQL functions available in the Generic Datasource Management System (GDMS) of OrbisGIS [9]. Among other functions, this open-source library facilitates the production of noise maps.

First, the user manipulates roads, buildings and non-geographical data with the native tools of OrbisGIS. The user can then independently call noise modules from the same graphical user interface in order to :

→ convert traffic flow data into noise level data

- compute the emission noise spectrum for a specified type of sound source
- compute the propagation of sound from the sources to the receivers

Second, built-in OrbisGIS functions are used for post-processing in order to :

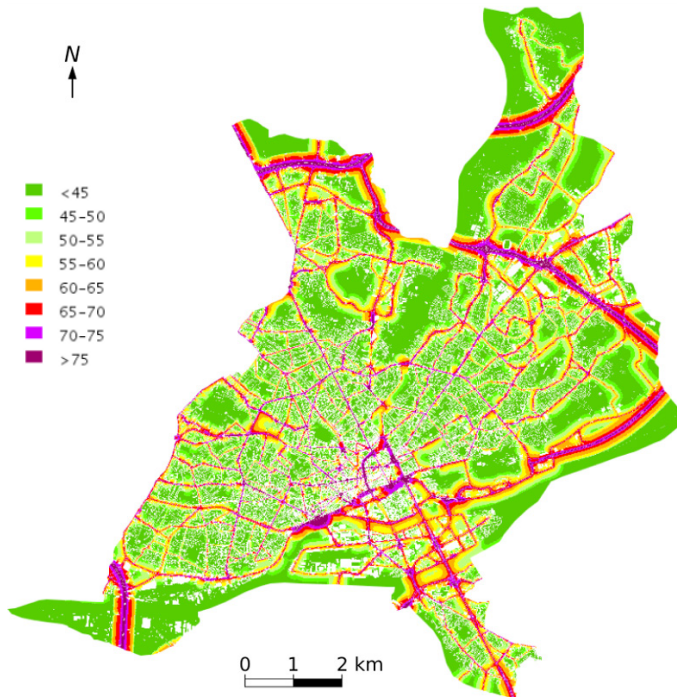
- interpolate the receivers' levels to construct a fully-customizable noise map
- combine the computation results in numerous ways
- establish a spatial statistical study of the results in association with external data
- publish intermediate or final results as files and/or over networks services

3. Case study

In the framework of the Eval-PDU project, several scenarios have been considered, not only for their impact on noise, but also for several other environmental (e.g., air pollution, energy use) and socio-economic (e.g., property values, behavioral changes, etc.) consequences. These scenarios are expected to express changes in citizens' displacement habits in comparison with the situation in 2008 due to variations in energy price, urban sprawling, and local and national economic changes. For our purposes, we only consider scenarios relating to noise impact :

1. a 25% decrease in road trac
2. a 20% increase in all trac (passenger cars, public transportation, etc.)
3. the doubling of gas prices
4. an increase in public transportation Nantes Metropole represents 24 communes and a surface area of 523 km²

Due to the large size of this study area, we restricted our attention to the city of Nantes (65 km²) only. The traffic data we used are built on top of geographical data using BD Topo, from the French Geographic Institute IGN⁴. The noise map uses the same database. Buildings represented as simple 2D polygons. In total, the database comprises 80,000 objects.

**FIGURE 4**

LDEN noise map of the commune of Nantes

As an illustration, the noise map of Nantes, shown in **Figure 4**, requires considering 345 million source-receiver couples for a total of 1.7 million calculations, with a computational time of approximately seven hours. This included the calculation of the direct and first-order reflected and diffracted fields, for a given time period and one octave band on a 2.13 GHz dual-core PC running Linux. We used the noise map to estimate the residents' exposure to noise in an attempt to understand the impacts of the various scenarios. We estimated this exposure by determining the number of residential units for each building and the number of residents per unit [8]. We then computed the maximum noise level at a distance of 1 meter in front of each facades. The number of residents exposed to an LDEN sound level greater than 68 dB can then be incorporated into more useful spatial objects such as commune boundaries or French statistical population units (INSEE).

Our results show that the differences among the scenarios we considered are marginal in terms of the affected populations. In effect, the number of residents exposed to a sound level greater than the threshold of 68 dB varies little depending on the scenario considered, in the range of 2,974 to 5,204. Nevertheless, we observe that even though the effects may be weak, they are precisely located.

[1] W. PROBST, *Methods for the calculation of sound propagation*, Proceeding DAGA, DEGA (2008).

[2] W. PROBST, *Multithreading, parallel computing and 64-bit mapping software-advanced techniques for large scale noise mapping*, DEGA (2008).

[3] R. REIJNEN, R. FOPPEN, G. VEENBAAS, *Disturbance by track of breeding birds: evaluation of the effects and considerations in planning and managing road corridors*, Biodiversity and Conservation 6(4), 567-581 (1997).

[4] F. BESNARD, J. DEFRANCE, M. BERENGIER, G. DUTILLEUX, F. JUNKER, D. ECOTIERE, E. LE DUC, M. BAULAC, B. BONHOMME, J.-P. DEPARIS, B. GAUVREAU, V. GUIZARD, H. LEFEVRE, V. STEIMER, D. VAN MAERCKE, V. ZOUBO, *Road noise prediction: 2 - Noise propagation computation method including meteorological effects (NMPB 2008)*, SETRA (2009).

[5] M. DREHER, G. DUTILLEUX, F. JUNKER, *Optimized 3D ray tracing algorithm for environmental acoustic studies*, Acoustics 2012, Nantes Conference (2012).

POSTERS

Free Software and Open Source in Education: Geoinformatics at the CTU in Prague

Author

Martin Landa, Czech Technical University in Prague, *Czech Republic*

KEYWORDS : Education, Free Software and Open Source, GIS, Remote Sensing, Programming

Free Software and Open Source (FOSS) plays an important role at Faculty of Civil Engineering, Czech Technical University (CTU) in Prague, study branch Geodesy, Cartography and Geoinformatics. There are several courses where the students learn how to use different Free Software and Open Source tools. These practical courses cover basic topics as introduction into database management systems, programming, and GIS-oriented subjects. Thanks to Professor Cepek Free Software and Open Source tools have been used at the CTU in Prague, study branch Geodesy, Cartography and Geoinformatics for several years.

In the first semester the students learn how to effectively use GNU/Linux operating system including some advanced topics like scripting in Bash or UNIX administrator-related issues. After introduction into GNU/Linux OS the students learn basics of SQL (Structured Query Language) which is a standard language for accessing object-relational databases. The assignments for Introduction into Database Systems course is developed with PostgreSQL as a widely used open-source object-relational database management system. For course evaluation is used GNU SQLTutor as a web based interactive tutorial of SQL. GNU SQLTutor [1] has been developed at the CTU in Prague by Professor Cepek in 2007. Later in 2010 have been added several geospatial datasets for interactive learning of spatial SQL [2] based on PostGIS geodatabase. In the third semester the students learn basics

of programming in C++. In this case the assignments were developed with QT framework. The basic programming course is focused on developing simple graphical-based applications using QT graphical library and QT framework for C++ programmers. In other semesters the students have also possibility to learn programming in Java using open-source development environment called Eclipse. The students also learn other FOSS tools as Octave for computations, R in statistics, LaTeX for writing technical reports or GNU Gama which is dedicated to adjustment of geodetic networks. The GNU Gama [3] is a project founded by Professor Cepek, and the most of contributors to this project are former or current students from the CTU in Prague, study branch Geodesy, Cartography and Geoinformatics. In the fifth semester the students work on semester project, the goal is to develop a web-based CGI application written in C++ using QT framework and light-weight database SQLite.

Similarly also GIS-related courses are focused on usage of Free Software and Open Source tools. Beside commonly widespread proprietary GIS platform provided by Esri company (Esri ArcGIS) the students learn about Free Software and Open Source for Geoinformatics (FOSS4G) in general - from community-driven development to the most important FOSS4G desktop and web-based projects. As we noted above PostgreSQL database system is used as a platform for teaching the Geoinformatics students at the CTU basics of SQL. PostGIS as a program that supports for geographic objects to the PostgreSQL object-relational database, was a natural choice when choosing platform for teaching the students how to store, manipulate and analyze geospatial data in object-relational database management systems. In other words, the assignments for Introduction into Spatial Data Processing course were mainly developed with PostGIS, some lessons are also dedicated to the SpatiaLite database as light-weight solution based on SQLite database.

The heart of FOSS4G in education at the CTU in Prague, study branch Geodesy, Cartography and Geoinformatics is Free Software in Geographic Information Systems (GIS) course [4]. Within this course, the students learn about the FOSS4G environment in general, including a community-driven development, or OSGeo role. In practical lessons the students learn how to use FOSS4G tools when manipulating, storing or analyzing geospatial data.

The assignments were developed with mostly used open-source GIS desktop applications like GRASS GIS, QGIS or gvSig. The students learn basics about development of web-based mapping applications using MapServer, and javascript library OpenLayers. The students also enhance their knowledge in programming or scripting, the assignments contain programming for GDAL/OGR library in C++ or Python, specific Python GIS-oriented libraries, introduction into writing C++ or Python plugins for QGIS or scripting in Python for GRASS GIS. One practical lesson is also dedicated to OpenStreetMap project including active mapping near the university campus. The students have also opportunity to borrow handy GPS at the university to do mapping for OpenStreetMap in their free time. The Free Software GIS course is also available in English to make it accessible for international Erasmus students.

Since 2005 the assignments for Remote Sensing course in the Geoinformatics program have been developed with GRASS GIS. The assignments are in Czech only, and available online [5]. GRASS GIS project is one of the most popular GIS and Remote Sensing tools world-wide. Beside GRASS GIS also Orfeo Toolbox or OSSIM are used for several tasks. The students have opportunity to report bugs in GRASS software or wishes directly to the teacher who is actively contributing to the GRASS GIS project since 2005. Smaller bugs are fixed or wishes implemented for the next lessons which makes the students happy and as a result interested in open-source software development.

Thanks to the fact that several courses at the CTU in Prague, study branch Geodesy, Cartography and Geoinformatics are based on usage of Free Software and Open Source, some of the Geoinformatics students from CTU became interested in contributing to the open-source projects, not only as developers (e.g. currently there are three our students which are actively contributing as developers to GRASS GIS project), but also as testers, bug-hunters, translators or power users. Since 2008 several Geoinformatics students from CTU participated in Google Summer of Code (GSoC) program and developed a new GUI in GRASS GIS for visualization of geospatial data in 3D (wxNviz, participation in 2008, 2010, and 2012) [6] or GUI front-end for GRASS vector network analysis tools (wxVNet, participation in 2012, [7]). Other GRASS-related projects which were developed by the Geoinformatics

students from CTU is wxGUI Cartographic Composer [8], wxGUI tool for supervised classification [9] or the new version of r.in.wms module in Python for downloading WMS (Web Map Service) data in GRASS GIS [10]. Some of the Geoinformatics students from CTU were also involved in QGIS development, recently has been developed by the students QGIS plugin for Czech cadastral data [11] or Workflow builder for QGIS Sextante project. The important point is that the students learn about different environments. The open society, open access to the knowledge and information, namely Free Software Foundation, projects which are based on open sharing of information like Wikipedia or OpenStreetMap play very important role nowadays. The students should simply learn about such aspects at the university.

[1] ČEPEK A. ET AL. 2007. GNU SQLTutor. Retrieved September 1, 2012 from <http://sqltutor.fsv.cvut.cz/cgi-bin/sqltutor>

[2] LANDA M., ČEPEK A. 2010. Learning PostGIS using Sqltutor. Retrieved August 16, 2012 from <http://geo.fsv.cvut.cz/~landa/publications/2010/foss4g-2010/paper/landa-cepek-sqltutor.pdf>

[3] ČEPEK A. ET AL. 2007. GNU Gama. Retrieved August 9, 2012 from <http://www.gnu.org/software/gama/>

[4] LANDA M. 2012. Free Software in GIS course at the CTU in Prague. Retrieved August 11, 2012 from http://geo.fsv.cvut.cz/gwiki/153YFSG_Free_software_GIS/en

[5] LANDA M. 2012. Remote sensing course at the CTU in Prague. Retrieved August 17, 2012 from <http://geo.fsv.cvut.cz/gwiki/153YZOD>

[6] LANDA M., KRATOCHVÍLOVÁ A. 2011. wxNviz Google Summer of Code project. Retrieved September 9, 2012 from <http://grass.osgeo.org/wiki/WxNviz>

[7] TUREK Š. 2012. wxVNet Google Summer of Code project. Retrieved September 8, 2012 from http://grass.osgeo.org/wiki/GRASS_GSoC_2012_WxGUI_front_end_for_vector_analysis_modules

[8] KRATOCHVÍLOVÁ A. 2012 WxGUI Cartographic Composer. Retrieved August 7, 2012 from http://grass.osgeo.org/wiki/WxGUI_Cartographic_Composer

[9] KRATOCHVÍLOVÁ A. PETRÁŠ V. 2011. wxGUI tool for supervised classification of imagery data. Retrieved August 9, 2012 from <http://grass.osgeo.org/wiki/WxIClass>

[10] TUREK Š. 2012. GRASS r.in.wms module. Retrieved August 15, 2012 from http://grass.osgeo.org/grass70/manuals/html70_user/r.in.wms.html

[11] KRATOCHVÍLOVÁ A., PETRÁŠ V. 2012. QGIS plugin for Czech Cadastral Data. Retrieved August 17, 2012 from http://grass.fsv.cvut.cz/gwiki/QGIS_VFK_Plugin

Spatial patterns analysis of environmental data using R

Authors

- **Carmen Vega**, University of Lausanne, *Switzerland*
- **Jean Golay**, University of Lausanne, *Switzerland*
- **Marj Tonini**, University of Lausanne, *Switzerland*
- **Mikhail Kanevski**, University of Lausanne, *Switzerland*

KEYWORDS : point process, spatial patterns, forest fire occurrence, K-function, Morisita index, fractal measures, multifractal, box-counting

Many environmental phenomena can be study as stochastic point processes where events are represented by their spatial locations (X , Y coordinates) within a specified geographical region. In this regard, taking into account the spatial characteristics of the environmental data, exploratory spatial data analysis methods are used to discover patterns of spatial association (spatial clustering). Thus, clustering analysis can reveal information about the patterns of the underlying process and their relationship with the phenomenon under study.

The present paper presents some exploratory spatial data analysis tools implemented in R software to assess the spatial patterns of the forest fire distribution in Switzerland. In this research we considered 2,402 georeferenced forest fires ignition-points occurring from 1969 to 2008 in Canton Ticino. This Canton is located in the Southern Swiss Alps and it is the most fire-prone region in Switzerland. The applied clustering measures are Morisita index, fractal and multifractal dimensions (box-counting) and Ripley's K -function. These algorithms enable a global spatial structural

analysis describing the spatial degree of clustering of a point pattern [1] and defining whether the observed events occur randomly, in clusters or in some regular pattern [2].

The spatial variability of forest fires is a very complex process which is conditioned by an intermixture of human, topographic, meteorological and vegetation factors. To compute measures of clustering in complex-shape regions the concept of validity domain is applied to restrict the spatial dimensionality of the phenomenon on the mapping space. Within the validity domain is possible to generate spatially randomly distributed events which structure properties are well known. These properties can be compared to the real phenomena properties [1, 3] and the deviation between these measures (computed using real data and based on randomly generated patterns) can quantify the real clustering.

Each measure is described and executed for the raw data (forest fires in Canton Ticino database) and results are compared to the ones obtained from the reference patterns generated under the null hypothesis of spatial randomness embedded in its validity domain. This comparison enables estimating the degree of the deviation of the real data from the random patterns. It is shown that the relative results are very different from measures computed in regular geometrical spaces (usually considered in the literature) because they include empty spaces.

Computations were carried out using R free software for statistical computing and graphics [4]. R is a free software environment integrating facilities for data manipulation, calculation and graphical display. The R base can be extended via packages available through the Comprehensive R Archive Network (CRAN) which covers a very wide range of modern statistics. More specifically, the spatial point pattern analyses considered in the present study were supported by the spatstat package [5] and customised functions.

Morisita index

Morisita Index is a statistical index used to characterise quantitatively the clustering of point processes, in this case the forest fire events (ignition points). Calculation of Morisita Index consists on dividing the study area

into identical Q quadrats of size D and counting the number of events n_i falling within each single i [1]. Morisita index is computed as follows: $I_D = Q ((S n_i (n_i - 1)) / N (N - 1))$, where N is the total number of points and I_D is Morisita index for a chosen quadrats of size D . In other words, this index measures how many times more likely it is to randomly select two points belonging to the same quadrat than it would be if the points were homogenously distributed [6]. Morisita index is first calculated for a relatively small quadrat's size which is then increased until it reaches a chosen value. If the studied point pattern is homogenously distributed on the studied area, every computed I_D fluctuates around the value of 1; while, if the points are clustered, the empty quadrats at small scales will increase the value of the index [1]. Morisita index computed for the forest fire dataset in Canton Ticino displays values above the unitary value for all considered quadrat's sizes which highlights the clustering of the fire distribution.

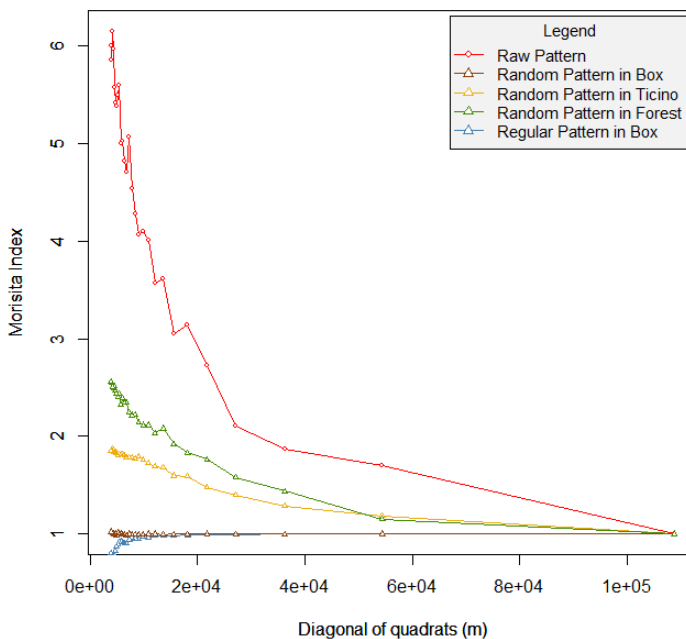


FIGURE 1

Box-counting

The box-counting method consists on superimpose a regular grid of boxes of length d on the region under study and count the number of boxes $N(d)$ necessary to cover the whole dataset. This procedure is repeated using different values of d . The algorithm goes on until a minimum diameter d is reached. The number of occupied boxes increases with decreasing box size, leading to the following power-law relationship: $N(d) = d^{-df_{box}}$, where df_{box} is the fractal dimension measured with the box-counting method [3]. The fractal dimension is estimated as the slope of the linear regression fitting the data of the plot which relates $\log(N(d))$ to $\log(d)$. If the observed events are homogeneously/randomly distributed in the studied area the number of boxes of diameter d necessary to cover the whole dataset decreases as d^2 . Consequently, the box-counting method enables the detection of clustering as a departure from a random situation [3]. The box-counting method diagram for the forest fires in Canton Ticino resulted in a df_{box} less than 2 indicating a high degree of clustering of the events distribution.

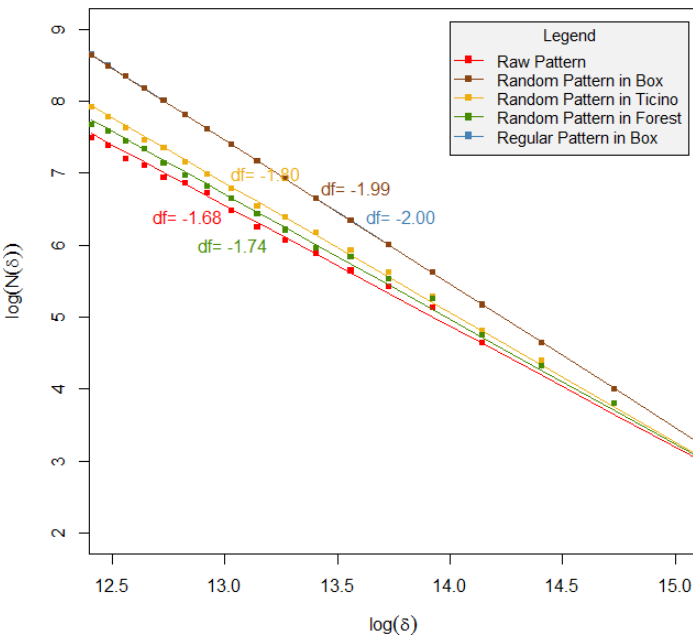


FIGURE 2

Multifractal characterisation

To describe the multi-scaling spatial structure of the forest fires distribution in Canton Ticino, a set of generalized q -dimensions, Dq , are measured through a generalization of the box-counting method. The generalized dimensions are defined in terms of the Rényi information of the q th order moment of the probability distribution. Let $N(d)$ be the number of non-overlapping boxes of equal size d needed to cover the fractal (observed events) and $p_i(d)$ the mass probability function in the i th box. The generalized dimension Dq is computed through the parameter q by [7]: $Dq = (1 / (1 - q)) \cdot \lim_{d \rightarrow 0} (\log (S (p_i(d)^q)) / \log (1 / d))$. The Dq diagram is obtained by calculating the generalized dimension varying q . When $q=0$, all nonempty boxes are equally weighted and Dq is equivalent to the box-counting dimension. For $q>0$, the mass within the boxes gradually gains more importance in the box contribution to the entropy value; therefore, the larger the mass values are in a box, the higher the weight of the box. Thus, Dq indicates the scaling of over-

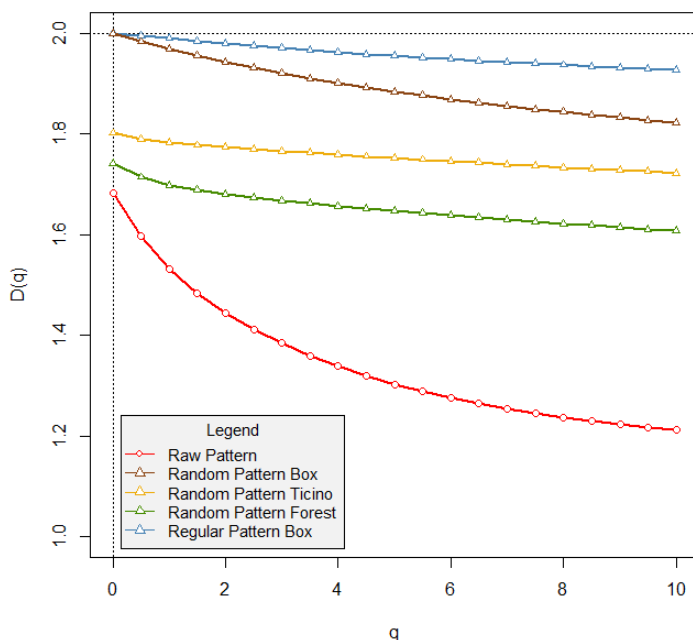


FIGURE 3

dense regions and strong singularities [7, 8]. The Dq for each q th moment exhibits a non-linear signature revealing the multifractal behaviour of the forest fires distribution in Canton Ticino. For the forest fire real data, Dq declines faster than for the generated random patterns, and the departure from these patterns reveals the highly clustering of the events distribution.

Ripley's K-function

The Ripley's K -function describes how the interaction between events varies through the space. For a spatial point process, the function is defined as: $K(r) = E(\times) / Intensity$, where $E(\times)$ denotes the expected number of further events within a distance r of an arbitrary event [2, 9]. The analysis tests for complete spatial randomness (CSR) which implies no interactions among points; that is, the probability of the occurrence of fire at any point is independent of each other fire event. Subsequently, departure from CSR indicates clustering (aggregation) or dispersion (regularity or inhibition) of fire occurrences.

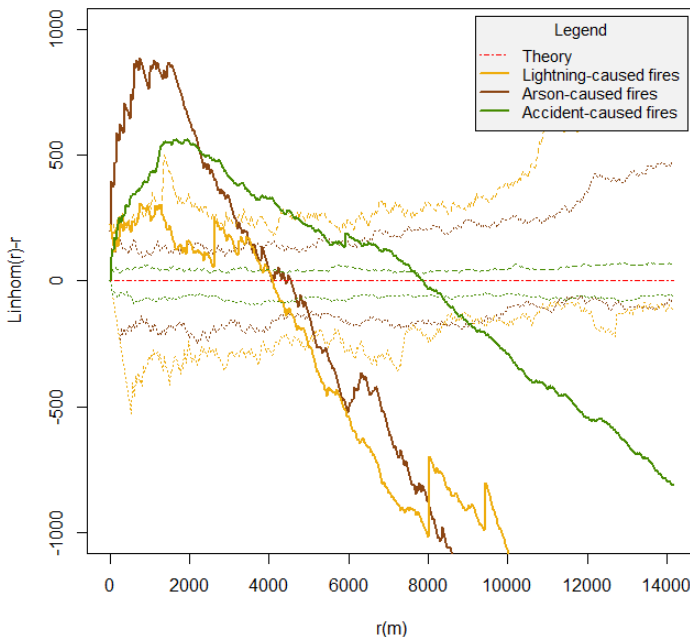


FIGURE 4

This is evaluated against a confidence interval constructed by performing n simulations of the events under the CSR hypothesis. This nonparametric spatial statistics tool can afterwards support nearest-neighbour analyses to detect areas with higher fire incidences. Results from this function revealed that human-caused fires are not randomly distributed presenting a clustering at scales up to 8km with a maximum degree of clustering at 2km.

Conclusions

Exploratory spatial data analysis tools were introduced to characterize the degree of clustering of the forest fire events in Canton Ticino. The concept of a validity domain was implemented. Each measure detected a quite higher degree of clustering of the forest fire network in the study area at every scale. This work demonstrates that in the case of complex regions relative values between measures are more important than absolute values. Further development of the research will consist in the custom implementation of the algorithms in the *R* environment to be functional for the analysis of multi-dimensional and multivariate point patterns.

[1] KANEVSKI M., MAIGNAN M. (2004). *Analysis and modelling of spatial environmental data*. Lausanne, EPFL Press.

[2] RIPLEY B.D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39 (2): 179-212.

[3] TUIAD., KANEVSKI M. (2008). *Environmental monitoring network characterization and clustering*. In "Advanced mapping of environmental data. Geostatistics, machine learning and bayesian maximum entropy". Kanevski M. (Editor), iSTE and Wiley, 19-46.

[4] R DEVELOPMENT CORE TEAM. (2012). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org/>.

- [5] BADDELEY A., TURNER R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, vol. 12 (6): 1-42.
- [6] HURLBERT S.H. (1990). Spatial Distribution of the Montane Unicorn. *Oikos*, vol. 58: 257-271.
- [7] SEURONT L. (2010). *Fractals and multifractals in ecology and aquatic science*. Boca Raton, CRC Press.
- [8] ADJALI I., APPLEY S. (2001). *The multifractal structure of the human population distribution*. In *Modelling scale in geographical information science*, edited by Tate N.J. and Atkinson P.M., West Sussex, John Wiley & Sons Ltd, 69-85.
- [9] HERING A.S., BELL C.L., GENTON M.G. (2009). Modeling spatio-temporal wildfire ignition point patterns. *Environmental and Ecological Statistics*, vol. 16: 225-250.

Design and Creating a Two-way Public Participation Geographical Information System Platform: An Open Source Architecture

Authors

- **Antonio Jose Fernandes da Silva**, Nova School of Statistic and Information Management (Instituto Superior de Estatística e Gestão de Informação - Universidade Nova de Lisboa), *Portugal*
- **Jorge Gustavo Rocha**, University of Minho Department of Informatics, *Portugal*

KEYWORDS : PPGIS, Open Government, Social Networks, Geospatial Crowdsourcing, Open source

Introduction

Until recently, government data made its way to the Internet mainly through raw data gathered by the work of civil servants, and offered these data to citizens seeking insight into many subjects. But a new, more dynamic approach is emerging – one that enlists private actors as allies in making government information and useful online.

Public Participation Geographic Information Systems (PPGIS) platforms are viewed in almost all democratic countries as an important tool in the management of territory, inviting ordinary people to give their opinion on relevant issues to their lives and bringing together citizens and government (central or local) to the center of decision, in order to facilitate, speed up and improve government decision-making processes.

PPGIS research is entering through its second decade. There has been much that has been accomplished, but there is much to do in regard to people's contributions to community well-being [5]. PPGIS projects are though still limited in their ability to communicate, organize, and reflect user participation. Current public participation geographic information systems (PPGIS) literature approaches geospatial collaboration from an empowerment or mobilization perspective [1, 2]. Recent PPGIS mapping initiatives seek to use GIS to incorporate information held by local communities into the planning process, to address concerns articulated by community participants and groups, to reduce inequalities in public access to information and technology, and to develop and make spatial information more adaptable for community use [7].

In a 1998 specialist meeting on "Empowerment, Marginalization and Public Participation GIS" Craig et al established some basic assumptions related to what a PPGIS should contain and look like [2]. These are:

- equal access to data and information is a key component of a Web-based PPGIS
- PPGIS should have the capability to empower the community and its members by providing the necessary data and information
- establishing and maintaining community trust is key for people working with PPGIS

Despite these basic assumptions, until recently, traditional PPGIS applications have been governed by an almost legal obligation imposed by government on citizens to participate by providing input on various matters of public interest, involving citizens on the one hand and government on the other.

Almost ten years later from the specialist meeting at the NCGIA, Elwood (2008) also stated that data access is an important precursor to citizen effective participation in planning and policy making, though data access alone does not guarantee an active and influential role.

A new paradigm of participation assumes collaboration of all interested parties. Collaborative participation can help build civic capacity at least among community leaders and they can spread it to their circles of associates. This capacity in turn has the potential to create a more intelligent society, better

able to adapt quickly to changes in the conditions and more competent to address controversial, difficult issues [4].

Recently, has emerged a new paradigm of PPGIS research, two-way PPGIS platforms who are characterized by allowing a bidirectional communication and an equal access to information between citizens and government, as opposed to traditional platforms that only encourage the participation in a one-way, from citizen to the government. Despite this, there aren't yet many two-way PPGIS platforms that accomplished the purpose of an equal access to information.

Citizens should be able to see what their impact is because of the transparency of the dialogue and openness of the conclusions, on supply-driven and pragmatic approaches to engage the public in applications of GIS with the goals of improving the transparency of and influencing government policy [4, 6]. Both citizens and government become providers and recipients of information. Such collaboration takes place in design groups and in Internet systems where users are actively engaged in design process.

Collaborative participation thus dissolves many dilemmas, for example, there is no need for citizens or planners to choose between the collective and individual interest. In these dialogues the effort to meet individual interests produces a collective interest. In collaborative participation, interdependencies are uncovered and participants can discover how all may benefit from improving a resource. Participants - public agencies, powerful private interests, and common citizens - are treated equally within the discussions. In these processes learning takes place, and sometimes conflicts are resolved and innovations emerge [4].

Many authors believe that practice will increasingly be defined by the collaborative model because it better serves the purposes of participation. These methods allow decision makers to learn more accurately about preferences because participants are more representative and have more opportunity to provide thoughtful, informed input than in the standard

required methods [4]. In order to achieve a successful two-way PPGIS platform, it is intended with this work to implement concepts of Social Networks and Geospatial Crowdsourcing in the design of an open source two-way PPGIS platforms.

Contextualization

In a context of a collaboration with the Agency for the Public Services Reform (AMA), we were challenged to connect the world of PPGIS and government, by completely redesign the existing platform “A minha rua1” (“My street”). “My street” is a service that is already running on the web, allowing citizens to report problems in their neighborhood. All reports are routed to the corresponding local or central authority in charge. But since some requirements were not initially identified, technological limitations are present, and the actual participation is low (with about 4 reported situations per day, for a population of 10 million).

Designing a Two-Way PPGIS Platform

The question of an successful two-way PPGIS platform is this: how can we use technology to make it into a better platform? We believe the platforms that are the most successful in these days are those that are the most open. By designing simple and open systems it is given an opportunity to others to build on, reuse and extend. Open source software projects like Linux and open systems like the Internet work not because there’s a central board of approval making sure that all the pieces fit together but because the original designers of the system laid down clear rules for cooperation and interoperability.

Social Media

We argue that two fundamental changes must exist in new PPGIS platforms: there is a shift from hierarchies to increased equal rights platforms; improved communication, more transparency, and bi-directionality.

The role of the authority in former PPGIS platforms was really an authoritarian role: having all the power and only partly knowing and controlling the entire

platform. This is completely different from the crowd source platforms we know to be successful. So, one fundamental change is to diminish hierarchies and prevent people from hiding themselves behind the institution.

The second major conceptual design issue is related to transparency and communication. While former platforms use mechanisms to prevent citizens from seeing each other's participation, we aim to enable people to see the participation of others, which is a fundamental feature in social media.

If citizens are requested to participate, the administration must use the same platform to communicate with them. Not only to provide feedback, but also to publish useful information for the citizen.

The very first paragraph of Wikipedia's "social media" definition says: "Social media includes web-based and mobile technologies used to turn communication into interactive dialogue". Since we definitely want to move from the traditional governmental monologue (traditional PPGIS platforms) to interactive dialogue, social media seems to promise that, while also complying with our other requirements as well.

We decided to move from a traditional web application where citizens can report problems, to a social network where dialogues occur between citizens and public servants. This dialogue is open. The problems reported are open, and the feedback is also open as well. Public servants not only answer problems reported by citizens, but also will publish information that might affect citizens daily life.

The existing tools, such as Twitter, YouTube or Facebook are public, hosted social networks for individuals to create user accounts. Since we want communication around people's neighborhood problems, we need to develop our own social network tool to integrate, among other smaller things, the concept of 'problem' and 'neighborhood'.

Instead of developing such a social network from scratch, we select one of the many social engines available. We needed an engine that already has the typical social network features implemented, such as user registration, user profile management, user groups, friendship support, messaging, discussions, etc. We chose Drupal because it has all such features already implemented and it is open source. Because it is open source the code of the platform it will be free and it will be possible for common users to extend and take fully advantage from the platform and the data.

To extend the social engine to support the “My Street” application, we started by implementing the notion of ‘problem’ that can be reported by the citizen. To submit a problem, the user publishes a slightly different post, since the problem’s location and its type must be specified. The location can be pinpointed on the map. The type of problem is chosen from a closed domain. After choosing the problem type, the user can add more information (and media) concerning the problem being reported. The reported problem is displayed like any other user’s post. Other users can see the reported problem and can say if the problem also affects them, as well as post additional media about it. Another extension of the social network engine is related to the spatial component. In the “My Street” application, the location is very important. It is important to know where the problems are, but it is also important to create relationships and engage communities around their neighborhood. In the user profile, users can add one or more locations, like their place of work and their residence. Accordingly, users can define more than one neighborhood they care about. These locations are used to filter problems for those areas related to the user. These areas are also used to filter the information published by the authorities. As soon as users log on, they can see all the recent information about their friends and group activities, as well as things that have been reported and fixed in their neighborhood(s). Location is a key component to implement this thematic social network.

Finally, parallel to the core social network back-office tools must exist to keep track of all the problems being reported. Internally, as soon as the problem is reported, this information is routed to the corresponding municipality and assigned to a specific public servant accordingly to its type. We need to know what problems are still pending solution, their duration, which problems have

been partially addressed, how long it takes for their solution, etc. Various statistics are collected and the performance of each municipality is openly published.

Examples of features that we want to include are:

- Advanced search
- RSS feeds
- Mashups with other data sources
- Discussion forums and wikis
- Visualization
- Automated content and topic analysis
- Collaborative filtering and crowdsourced analysis

Outlook

In this paper, a new paradigm of PPGIS platforms are translated in concrete design solutions to be developed on top of an open source platform, already supporting typical social network features, like friendship, messaging, etc. The goal in the short term is to improve the relation and consequently the participation between the citizens and the authorities. While the conceptual issues will remain valid for a larger period, the concrete design issues can be constantly improved. For that purpose, the platform itself is open source and it is designed from the beginning to allow its enhancement and the development of extensions. While it might be difficult for the average citizen to improve its implementation, it would be great that public and private companies would like to enhance the platform with extensions that can be used to communicate with other stakeholders related to the public space: recycling companies, lighting companies, transit operators, etc.

Since the implementation is open source, porting this PPGIS platform for other environments like set-top boxes, smart tvs, mobile phones and tablets became easier and less expensive. Such porting costs can be supported by content providers, for example, to keep users connected on their platforms.

- [1] BAILEY, K. AND GROSSARDT, T. 2010. Toward Structured Public Involvement: Justice, Geography and Collaborative Geospatial/Geovisual Decision Support Systems. *Annals of the Association of American Geographers*. 100, 1 (Jan. 2010), 57-86.
- [2] CRAIG, W. ET AL. 1999. Empowerment , *Marginalization and Public Participation GIS - Report of Varenius Workshop*.
- [3] ELWOOD, S. 2008. Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*. 22, 1 (Jan. 2008), 71-90.
- [4] INNES, J.E. AND BOOHER, D.E. 2004. Reframing public participation: strategies for the 21st century. *Planning Theory & Practice*. 5, 4 (Dec. 2004), 419-436.
- [5] NYERGES, T. AND AGUIRRE, R.W. 2011. Public Participation in Analytic-Deliberative Decision Making: Evaluating a Large-Group Online Field Experiment. *Annals of the Association of American Geographers*. 101, 3 (Apr. 2011), 561-586.
- [6] SIEBER, R. 2008. Public Participation Geographic Information Systems: A Literature Review and Framework. April 2012 (2008), 37-41.
- [7] WOOD, J. 2005. How green is my valley?' Desktop geographic information systems as a community-based participatory mapping tool. *Area*. 37, 2 (2005), 159-170.

OSM indoor: moving forward

Authors

- **Jorge Rocha**, University of Minho, *Portugal*
- **Nair Alves**, University of Minho, *Portugal*

KEYWORDS : OSM, Indoor navigation, Indoor routing, VGI, 3D

Introduction

The OpenStreetMap (OSM) project started focused on mapping streets. But as many other successful innovations, it created new possibilities. Beyond streets, the community started to capture other information, to create more than just street maps. For example, initiatives like wheelmap.org enhance the OSM with tools and specific data for wheel drivers.

The OSM proved that Volunteered Geographic Information (VGI) [3] can be used in large scale projects. OSM became a school for neogeographers, where they learn how to produce high value user-generated geographic content. Many techniques and technologies were developed specifically to improve the quality of OSM, and to support the OSM mappers.

In this paper we aim to present a proposal to map indoor spaces in OSM. The goal is to map public indoor spaces, like universities, malls and airports. A public hospital, for example, can have many more people moving around than a small city. This is not the first or the only proposal around. Many people already suggested and developed some support to indoor spaces in OSM. This proposal tries to get the best of former initiatives to establish a common path to move forward.

We start by reviewing some important concepts, obviously starting by defining the meaning and the scope of indoor space and its representation. There are representations based on floor footprints and representations more based on the building model. Either can be more syntactical or semantic oriented. The distinction between indoor spaces and 3D models is necessarily discussed, since the distinction between them is somewhat blurred. We also revisit the commercial notion of 3D maps, used by companies like Google, Microsoft or Nokia in their web mapping solutions. These applications are 3D in the sense that they provide terrain elevation plus the 3D building volumes. There are no building models involved. Just the volume (the shell) of the building is modelled and portrayed with textures. For OSM, there are proposals to add this volumetric information to OSM [12], and Heidelberg OSM 3D [7, 13] is a good example of such proposal. More sophisticated approaches are also emerging [1, 2], using the entire building information model (BIM), instead of just its volume. In this scenario, the indoor space can be completely represented in the building information model. Although we strongly believe that is feasible to have BIM shared on OSM in the short time, we argue that OSM indoor mapping still makes sense, and it will take some time to OSM indoor be completely computed from shared BIM. It is worth to remember that OSM indoor has some advantages: it works pretty well on 2D, people are already trained to read floor plans, the existent editors like Potlatch and JOSM can be used to draw floor plans with minimal changes and the existent portrayal technology does not require any changes, since multiple floors can be drawn on top of the building bounds. 3D building models are more powerful and better represent the building, since they have more information. But the complexity of 3D BIM is its major disadvantage. All the increased complexity in modelling, representation and manipulation can not be easily transformed in an incredible better visualization. 3D navigation is complex for ordinary users.

Users are not trained to navigate and understand 3D models. For that reason, we argue that 2D visualizations of indoor spaces are necessary and are a major enhancement over the existent OSM data. It will take more time to change to overall OSM work flow to support 3D BIM. We also argue that both models can coexist, and we can have indoor support and 3D BIM data

simultaneously. The indoor can be helpful to improve the 3D models. Finally, indoor routing and indoor navigation are also reviewed, to highlight the differences and the relations between them. Any kind of routing can always be reduced to a graph representation, but current available algorithms rely on the distance to populate the graph. For indoor spaces, the same location can be used for an elevator that travels through several floors, for example. The graph resulting from the indoor space must be aware of the distance between nodes on different floors, but on almost the same location. Indoor navigation techniques are also reviewed. All techniques are grouped in Pedestrian Dead Reckoning (PDR), Radio Frequency based (WiFi, GPS+beacon, UWB, RFID, Bluetooth and NFC) and Image Analysis based [4].

OSM Indoor related initiatives

After reviewing the main concepts related to indoor spaces, we analyse three known proposals for OSM Indoor. The scientific literature is scarce about the subject. The only publications we know about are [5, 6]. Besides these publications, the best starting point for our research was the OSM Wiki, where mostly of the discussions take place, and where pointers to examples are posted.

The methodology adopted for our analysis was to review and compare each proposal against a list of characteristics, group in four categories:

- space representation
- supporting tools
- generation of the routing network (graph) and routing calculation
- interaction, visualization and exploration

We also surveyed some commercial solutions, mostly to evaluate some HCI issues, since we don't have access to the underlying implementation.

All the three initiatives are related to university buildings or campus: Heidelberg University, Germany [8, 9]; Alpen Adria Campus of the University of Klagenfurt, Austria [11]; and Gløshaugen campus of the Norwegian University of Science and Technology [10].

Proposal

The proposal presented for the OSM indoor support take advantage of experiments already mentioned. The goal is to point way a common way to handle indoor spaces in OSM. If we are able to generate some consensus about a common approach, many more mappers will help to improve the tools and share their knowledge about indoor space representation. The proposal covers the space representation, all necessary tools, and a new web interface, able to explore the indoor space and to provide the necessary controls to enable route calculation.

Space representation

This proposal is for indoor spaces, and not for 3D BIM. The indoor initiatives reviewed proved us that the current OSM data model is able to represent the indoor space. So, the impact on the overall OSM architecture and work flow is minimal, since the data is stored on the same database. There are no modifications needed on the current API. By the contrary, if we want to represent 3D BIM on OSM, we should move these to another database, with another data model and API. The OSM data model is so simple (nodes, ways and relations), that theoretically we can represent any building with that primitives. But the abstraction level of that representation is not suited. 3D BIM requires a separated database, with models and API that manipulate concepts more closed to the building semantics. The recent proposal [1] goes in that direction, since they propose a new parallel project, OpenBuildingModels for 3D BIM data.

Tagging

For OSM Indoor two major proposals are being discussed on the OSM wiki. One, from the University of Heidelberg, tries to represent more than the indoor space, and has more tags to represent the building details#. This proposal almost tries to represent the 3D building model using OSM tags. The other proposal, by User:Saerdnaer, is more simple#. We propose that this last proposal be used as the starting point. Some tags from the first proposal can be used, not for OSM indoor, but for OSM 3D. By other words, we should distinguished the minimal tags necessary for indoor from tags than can be used to better extrude 3D buildings (just the shell) from the OSM data.

Tools

Editors can be slightly improved to enhance the representation of indoor spaces. Basically, the existent editors can provide an interface where users can edit each level independently. Additionally, an OpenLayers based editor can be developed, using the simple editing primitives of OpenLayers. While this might not be the best solution, it can be an alternative.

Routing

The routing solution is as simple as possible, and only includes routes by foot or by wheel chairs. As we already said, the only specific requirement is related with the distance weight of distinct nodes at the same location. From our initial tests, the precision of locations is relevant, since locations are more closed. When generating the network graph, the entire precision of the locations must be considered, without any kind of rounding.

Web map

A new web interface is necessary to fully explore the indoor additional information. The new interface should be aware of existing indoor data. At certain zoom levels, if indoor data is available, an additional control is displayed. The number of floors will be used to display the floor switch. To calculate routes, two possibilities are provided. The origin and destination can be pointed on the map, or the user can select the explicit rooms or stores, if they were tagged on the data. The result route is calculated on the server side, and it is returned as WFS feature. Since the sources of this alternative site are open source, users can improve these features. One additional feature is to use such interface within other web sites, for example, in mall or airport web sites, where they can take advantage of the crowded source data and tools.

Conclusions and outlook

This discussion and distinction between OSM Indoor and OSM support for 3D BIM is very important to guarantee that we focus on OSM indoor without the temptation to include BIM features. In this paper we discussed the difference between both approaches and we argue that both models can co-exist in OSM.

Having stated that, the OSM indoor is a wise compromise, to enhance the OSM, without moving right now to fully support 3D BIM. That will take time, and OSM indoor will be around for a while, since users are more familiar with this kind of space representation.

For OSM indoor, we need to create some broad consensus about the tags, but it would be easier if we provide tools to support and explore that additional information. Together is this proposal, we started an open source implementation of all necessary tools to fully explore the indoor data, as described.

[1] UDEN, M., & ZIPF, A. (2012). OpenBuildingModels - Towards a Platform for Crowdsourcing Virtual 3D Cities, 1-17.

[2] NAVIGATION, I., & SCANNER, T. L. (2008). 3D Indoor Navigation : a Framework of Combining BIM with 3D GIS. Building, 1-10.

[3] GOODCHILD, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221. doi:10.1007/s10708-007-9111-y

[4] HAKAN KOYUNCU AND SHUANG HUA YANG. A Survey of Indoor Positioning and Object Locating Systems. *Journal of Computer Science*, 10(5):121{128, 2010.

[5] GOETZ, M., & ZIPF, A. (2010). Extending OpenStreetMap to Indoor Environments : Bringing Volunteered Geographic Information to the Next Level, (Hansen 2004).

[6] HALVORSEN, C. (2011). Campusguiden - En navigasjonstjeneste for innendørs bruk. Master's Thesis, Norwegian University.

[7] ZIPF, A. (2010). University of Heidelberg OGC Services Emergency Routing with OpenLS RouteService 3D Visualisation with W3D in the GEOSS Architecture Implementation Pilot – Phase 3. Response, 49(March).

[8] GISCIENCE GROUP. 2011. IndoorOSM - Mapping the Indoor World (2D). *Institute of Geography, University of Heidelberg*. Retrieved July 25, 2012 from: <http://indoorosm.uni-hd.de/>

[9] GISCIENCE GROUP. 2011. IndoorOSM - Mapping the Indoor World (3D). *Institute of Geography, University of Heidelberg*. Retrieved July 25, 2012 from: <http://indoorosm.uni-hd.de/3d/>

[10] HALVORSEN, C. 2011. Campus Guiden. *Gløshaugen campus of NTNU, Norwegian*. Retrieved July 30, 2012 from: <http://www.campusguiden.no/?lang=english>.

[11] GOMOGI. 2012. Campus-GIS. *Alpen Adria Universität Klagenfurt*. Retrieved July 27, 2012 from: <http://opengeo.org/gallery/klagenfurt/>

[12] OSM WIKI. 2012. Indoor Mapping. *OpenStreetMap Wiki*. Retrieved August 6, 2012 from: http://wiki.openstreetmap.org/wiki/Indoor_Mapping

[13] OSM WIKI. 2012. OSM-3D. *OpenStreetMap Wiki*. Retrieved July 26, 2012 from: <http://wiki.openstreetmap.org/wiki/OSM-3D>

Multimodal Planner: From Prototype to Production

Author

Francisco José Peñarrubia, SCOLAB Software Colaborativo,
Spain

KEYWORDS : planner, multimodal, smart city, bus, metro, bike sharing, open source, routing, public transport

Introduction

In 2010 we started to develop a new Multimodal Web Trip Planner for the city of Valencia. Those days, a new open source project was been developed (OpenTripPlanner [1]), and we decided to use it to achieve our goals: The planner should be able to calculate trips using buses, metro, bikes and/or walking.

In this paper we will present what problems we faced by using a software who was not finished yet, what decisions we took that lead to good (or bad) results, and some lessons learned for the future. We will also explain one innovative algorithm that permits to calculate routes using bike sharing service, mixing bus + metro + bike sharing, and the problems we had (and how we solved them) when the system went to production mode.

Next, we will present the final version of the web [2], remarking some of its more interesting features :

- Multimodal trip planner, offering several alternatives
- Real time information
- Multi language
- Printing capabilities
- Added information (timetables, points of interest, nearest points, etc)
- Integration with OGC Services (the background map is a WMS service, and the bus lines and points of interest and stops come from a WFS service)

Implementation

The project has been developed totally with Free Software. The server side is implemented using projects like Mapserver [3], Geoserver [4] and TileCache [5], following the recommendations of OGC (Open Geospatial Consortium [6]). Mapserver is used as a WMS server, and Geoserver implements the WFS services, publishing layers like bus lines, stops and Points of Interest. In order to speed up the rendering of the map, we used Tilecache to publish the tiles generated.

The data is stored in PostgreSQL + PostGIS [7] database, and the graph is generated from GTFS files [8] supplied by EMT (buses) and Metro. The core of the solution is the routing service, based in OpenTripPlanner project.. This project is made up of 3 principal subprojects. The first part is used to generate the graph file from GTFS files + streets. The second part is a routing service (planner) that reads the graph and receives the requests from the client. The third part is in the client side, and offers a user interface to allow specify the origin and destination, hour of the trip, and several parameters more. It uses 2 main JavaScript libraries to achieve a good looking and to support the most important internet browsers in the market (ExtJS [9] + OpenLayers[10]).

Also, to prepare the cartography and the data, we used gvSIG [11], and some new tools have been implemented in order to check topology and incorporate data. About our innovation, the bike sharing problem is a bit different than the rest of the transportation problems. It is similar, although there are some differences :

- Bus and metro gives uses information about timetables in a stop. In bike sharing, you don't wait in the bike station. If there are bikes, you just take the bike and go
- Bus and metro always follow the same path (there are defined lines) but in bike sharing, the user will use the streets network to reach his destination
- Also, there were some minor parameters to take into account (use the bike lines where possible, follow the sense of direction in the streets, etc

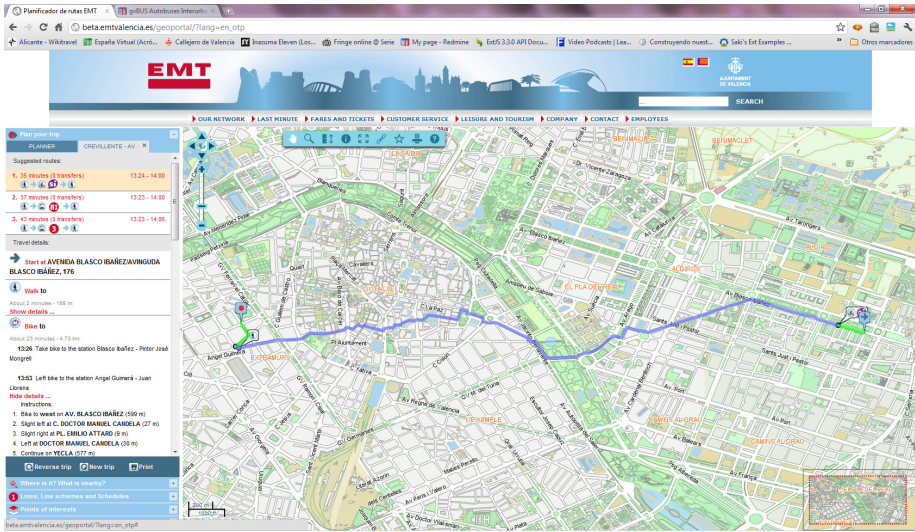


FIGURE 1

Bike sharing calculation

The approach to solve the bike sharing integration with the route planner is based on the idea of creating a virtual network where the bike sharing user (ValenBisi [12]) can go to a bike station, take the bike, and in this moment, the labeling of network nodes will run with the virtual network. This way, the minimal time in a node is different if the user arrives there walking or by bike.

Then, the user rides the bike to a close bike station of the destination and there, he left the bike in the station and start walking, maybe to arrive to the destination, or maybe, to take another transportation mode (bus or metro). This is the improvement, the calculation mixes bus, metro and/or bikes, allowing to mix every kind of transportation.

There was some difficult issues solved that it would be good to take care in the future. One was the requirement to keep compatibility with some old services in use. With this feature in mind, it was a good decision to use a common (and open model). Goggle Transit Feed System (GTFS) file was created to send

data to Google Transit, and this file is the start point to calculate the Graph object suitable to OpenTripPlanner routing service.

We faced also a lot of problems in client side programming because JavaScript behavior is different in Internet Explorer, Chrome, FireFox, etc. The library OpenLayers and ExtJS helped us a lot, but in the end, we expend very much time in debugging JavaScript.

And finally, when we put the production system running, we had problems with memory consumption and the server going down when we received many requests. To avoid this in the future, a good testing plan should be chosen, and is good to be prepared to increase RAM or processors in the server.

Also, is good to separate routing services from mapping services. In our case, the map requests (tile cache) were serviced by MapServer + TileCache, so, each time a new user enter the geoportal, lot of request (each tile for every layer) is requested, and also some WFS requests to get information about the stops in the visible extent. The Apache server was in pre-fork mode, due to some php limitations in other parts of the portal. This means that every single request will be attended by a new process, and this leads to more memory consumption. This was the main reason that put the server down when lot of users accessed the geoportal at the same time. The memory was used by Apache, and Tomcat (where the routing service is) couldn't work well. Luckily, all these problems were solved (and many more that are outside of this paper) and the geoportal now is up and in production mode.

Conclusions

Summing up, the experience has been very positive, and the client is very happy with the result. Each day, around 3000-4000 route calculations are done, and the user experience has been improved very much.

The advantages of this web compared to Google Transit, for example are :

- Better information and easy to maintain. One of the problems that EMT wanted to solve with this system was related to the delay between the GTFS delivery and the use of the new data from Google. Those days,

around 15 days was usual, so, it was impossible to reflect incidences in the web page. (Now, Google is testing in USA a new format that permits real time communication of incidences, but is still not ready in Europe)

- Real time information about arrivals
- Street data and tourist info directly maintain by the cartography service from Valencia's city council
- Better integration with the rest of the web page
- Support for ValenBisi information and really multimodal integration

In near future, a mobile version will be released (for iPhone and Android phones) and the new bike sharing algorithm may be used for other kind of sharing (car sharing, electric bikes, and so on).

[1] OpenTripPlanner wiki. GitHub. Retrieved August 31, 2012 from <https://github.com/openplans/OpenTripPlanner/wiki/>

[2] EMT. Empresa Municipal de Transportes geoportal page. Retrieved August 15 , 2012 from http://www.emtvalencia.es/geoportal/?lang=en_otp

[3] MapServer project homepage. Retrieved September 3, 2012 from <http://mapserver.org/>

[4] GARNET J. AND OTHERS. GeoServer web page. Retrieved September 3, 2012 from <http://geoserver.org/>

[5] TILECACHE CONTRIBUTORS. TileCache wiki. Retrieved September 3, 2012 from <http://tilecache.org/>

[6] OGC. Open Geospatial Consortium homepage. Retrieved September 3, 2012 from <http://www.opengeospatial.org/>

[7] Postgis homepage. Retrieved September 3, 2012 from <http://postgis.refractions.net/>

[8] GOOGLE. General Transit Feed Specification Reference. Retrieved September 3, 2012 from <https://developers.google.com/transit/gtfs/reference>

[9] SENCHA INC. ExtJS library web page (3.x). Retrieved September 3, 2012 from <http://www.sencha.com/products/extjs3/>

[10] OpenLayers wiki. Retrieved September 3, 2012 from <http://openlayers.org/>

[11] GVSIG ASSOCIATION. gvSIG project homepage. gvSIG Portal. Retrieved August 31, 2012 from <http://www.gvsig.org>

[12] JDCDECAUX ESPAÑA S.L.U. ValenBisi homepage. Retrieved September 3, 2012 from <http://www.valenbisi.es/>

The ESA BEAM Toolbox and Development Platform

Authors

- **Norman Fomferra**, Brockmann Consult GmbH, *Germany*
- **Daniel Odermatt**, Brockmann Consult GmbH, *Switzerland*
- **Carsten Brockmann**, Brockmann Consult, *Germany*
- **Peter Regner**, ESA ESRIN, *Germany*

KEYWORDS : Satellite remote sensing, Earth observation, European Space Agency, image processing toolbox, application development platform

The BEAM project [1, 2] was kicked-off 10 years ago, in June 2002, in response to a call by the European Space Agency (ESA). The objective was to provide a set of open source tools and application programming interfaces for scientific use. An open source policy was chosen in order to ensure technical transparency, foster the implementation and exchange of innovative algorithms, and avoid proprietary licensing restrictions. BEAM originally aimed at the exploitation of remotely sensed data products from Envisat's MERIS, AATSR and ASAR instruments.

After its first release, the BEAM toolbox was very well received by data users and also software developers. While being further developed over the years, 24 version updates have been released so far, turning the basic Envisat toolbox into the *BEAM Development Platform* for general Earth Observation data visualisation, analysis and processing. Due to its open and modular architecture and its extendibility through plug-ins BEAM comprises a large number of tools supporting a variety of Earth Observation sensors and data formats. Today, BEAM counts a few thousand users and is actively developed and maintained by Brockmann Consult. External collaborators contribute to the BEAM source code and external organisations provide additional

plug-ins. The most recent versions can be downloaded from <http://www.brockmann-consult.de/beam/>.

The main components of BEAM are :

- VISAT - An intuitive desktop application used for Earth observation data visualisation, analysing and processing
- A set of scientific data processors running either from the command-line or invoked by VISAT
- The command-line tool “gpt” (graph processing tool) is used to execute processing graphs made up of operators nodes developed using the BEAM graph processing framework (see Java application programming interface below)
- A data product converter tool allowing a user to convert raw data products to the BEAM-DIMAP standard format, to NetCDF, HDF, GeoTIFF, or RGB images
- The development platform, a rich Java™ application programming interface which provides ready-to-use components for remote sensing related application development and plug-in points for new BEAM extension modules. Besides a number of extension points such as product reader and writers, the BEAM application programming interface comprises the graph processing framework which is used to rapidly create raster data processors. The VISAT rich client platform is used to develop rich graphical user interface applications based on BEAM VISAT

The most important extension points provided by the BEAM Development Platform are:

- Product I/O application programming interface – allows to extend BEAM for new data product readers and writers
- Graph processing framework - This application programming interface allows to rapidly develop new data processors (processing “operators”) and to combine them to processing graphs
- VISAT rich client platform – Used to build efficient Earth observation imaging applications for the desktop

The following features are provided by the BEAM/VISAT desktop application:

- Very fast image display and navigation even of giga-pixel images
- Advanced layer management allows adding and manipulation of new overlays such as images of other bands, images from web map servers or ESRI™ shapefiles
- Rich region-of-interest definitions for statistics and various plots
- Easy bitmask definition and overlay
- Flexible band arithmetic using arbitrary mathematical expressions
- Accurate reprojection and ortho-rectification to all common map projections,
- Geo-coding and rectification using ground control points
- Store and restore the current session including all opened files, views and layers
- A standard set of scientific data processors which includes
 - ▶ Level 3 binning and mosaicing (all sensors)
 - ▶ Collocation (all sensors)
 - ▶ Expectation-minimization (EM) and K-means clustering, linear spectral unmixing (all sensors)
 - ▶ Radiance-to-reflectance, smile effect correction, cloud probability, SMAC atmospheric correction, case 2 water constituents (MERIS)
 - ▶ Sea surface temperature (AATSR)
 - ▶ Fluorescence line height, maximum chlorophyll index, normalized difference vegetation index (all sensors)
 - ▶ And many more

The supported sensors are MERIS, AATSR, ASAR of Envisat, ATSR and SAR of ERS, ALOS, AVNIR-2 of PRISM, MODIS of Aqua and Terra, CHRIS of PROBA, AVHRR of NOAA-KLM and MetOp, VGT of SPOT, TM of Landsat and many more available as plug-ins. Furthermore BEAM supports opening files of a number of formats commonly used in the modelling and remote sensing domain such as GeoTIFF, NetCDF CF and HDF-EOS.

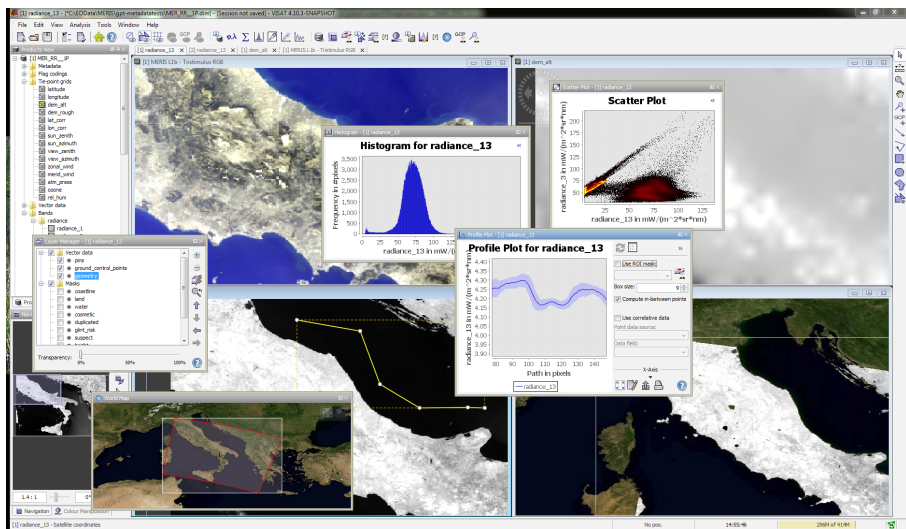


FIGURE 1
BEAM VISAT user interface

The BEAM development platform is the basis of a number of very efficient Earth observation data applications. Among those, the most prominent application is NEST, a specialisation of BEAM comprising many new readers, processors and graphical user interface extensions dedicated to the exploitation of synthetic aperture radar (SAR) data.

In the latest version 4.10 (May 2012), BEAM has been enhanced by a number of efficient and user-friendly tools supporting validation activities. A number of new features are resulting from a fruitful collaboration with the NASA Ocean Biology Processing Group (OBPG). OBPG is developing SeaDAS, a comprehensive image analysis package for the processing, display, analysis, and quality control of ocean color data. OBPG is currently replacing the SeaDAS 6 frontend using BEAM's VISAT rich client platform. The new SeaDAS 7 frontend will also offer a number of great new tools for SeaDAS users. All SeaDAS command-line processing tools (*l2gen*, *l3gen*, etc) will be callable from modern and intuitive user interfaces. The new SeaDAS 7 frontend will of course support all data products offered through the OPBG, namely MODIS, SeaWiFS, CZCS, OCTS and VIIRS.

Recently ESA has selected the BEAM development platform as basis for the Sentinel exploitation tools. The main objective of the development of Sentinel Exploitation Tools is to provide users a means to work with all standard data products generated by the sensors on board of the new Sentinel satellites which will be launched 2013/2014. This includes fast visualisation, comprehensive analysis and effective processing of the data. The requested way to achieve this is to extend the existing ESA toolboxes BEAM and NEST by dedicated reader plug-ins for the Sentinel user products. This way, the majority of the existing NEST (for the Sentinel 1 data products) and BEAM (for the Sentinel 2 and 3 data products) features are applicable to these data products.

In addition to a number of dedicated readers, the Sentinel 2 atmospheric correction processor for Sentinel 2's multi-spectral instrument (MSI) level 1 data will be integrated into the BEAM toolbox. Developing efficient tools for the Sentinel data products is a challenging task and is due to the greatly increased data volumes. Single Sentinel data products are magnitudes larger than their Envisat counterparts, and they do provide much more information, including increased spatial and spectral resolution coupled with the provision of per-pixel annotations such as geo-location, sensing time, measurement uncertainties and other quality information. The Sentinel exploitation tools will make available the following data products for a maximum of existing BEAM and NEST visualisation, analysis and processing tools:

- Sentinel 1 (SAR): level 0, level 1 single look complex, level 1 ground range detected, ocean level 2 products (wind, wave, currents)
- Sentinel 2 (MSI): level 1B, level 1C, level 2A (surface reflectance product)
- Sentinel 3 (OLCI): level 1B, water level 2, land level 2
- Sentinel 3 (SLSTR): level 1B, water level 2, land level 2; synergy level 1C, level 2; VGT P, S1, S10

[1] FOMFERRA, N., PETERS M., QUAUST, R., ZUEHLKE, M., DANNE, O., STORM, T., BROCKMANN, C., REGNER, P. ESA'S Toolboxes for optical Earth observation data: BEAM, CHRISBOX and the GLOBTOOLBOX. In *Proceedings of ENVISAT Symposium 2010* (Bergen, Norway, 2010).

[2] BROCKMANN, C., FOMFERRA, N., PETERS, M., ZÜHLKE, M., REGNER, P., DOERFFER, R. A programming environment for prototyping new algorithms for AATSR and MERIS – BEAM. In *Proceedings of ENVISAT Symposium 2007* (Frascati, Italy, 2007).

WPS tools to support geological and geomorphological mapping

Authors

- **Ivan Marchesini**, CNR IRPI, Perugia, *Italy*
- **Mauro Rossi**, CNR IRPI, Perugia, *Italy*
- **Massimiliano Alvioli**, CNR IRPI, Perugia, *Italy*
- **Michele Santangelo**, CNR IRPI, Perugia, *Italy*
- **Mauro Cardinali**, CNR IRPI, Perugia, *Italy*
- **Paola Reichenbach**, CNR IRPI, Perugia, *Italy*
- **Francesca Ardizzone**, CNR IRPI, Perugia, *Italy*
- **Federica Fiorucci**, CNR IRPI, Perugia, *Italy*
- **Vinicio Balducci**, CNR IRPI, Perugia, *Italy*
- **Alessandro Cesare Mondini**, CNR IRPI, Perugia, *Italy*
- **Fausto Guzzetti**, CNR IRPI, Perugia, *Italy*

KEYWORDS : WPS, geology, geomorphology, open source, SDI

The Open Geospatial Consortium (OGC) Web Service (WS) standards are technical documents developed for World Wide Web applications designed to share information, and facilitate the inter-operability of geospatial data. In particular, Web Coverage Service (WCS), Web Mapping Service (WMS), Web Feature Service (WFS), and Catalogue Service for the Web (CSW) are part of the Information Management Web Services Tier [17] to share metadata, and raster and vector thematic maps. An increasing number of Spatial Data Infrastructures (SDIs) is being developed and implemented to provide effective access to geospatial data exploiting OGC services.

Examples are available for disaster research and risk management [12] [15] [20], monitoring of glaciers [18], marine administration [25], e-Government [10], and the management of groundwater resources [11], coal mines [23], and generic environmental data [29].

The Web Processing Service (WPS) is a specific OGC standard for the design of interfaces to facilitate the publication of geospatial processes [13]. Implementation of WPSs remains less common than the implementation of WMS, WFS, and WCS, and is practiced chiefly by the Open Source (OS) community [27]. Multiple software can be used to deploy a WPS, including Deegree (<http://www.deegree.org>), 52°North (<http://52north.org/>), GeoServer (<http://geoserver.org/>), and the Zoo Project (<http://zoo-project.org/>) [5]. For uDig (<http://udig.refractor.net/>) and QGIS (<http://www.qgis.org/>) client GIS software, specific WPS plugins are available. GRASS GIS 7 offers a method to create a WPS, and can be integrated in a WPS server through the “wps-grass-bridge” framework (<http://code.google.com/p/wps-grass-bridge/>). Multiple WPS processes can be linked in a single processing chain [4], and can be run on a cloud or a grid infrastructure, for improved performance [6]. WPS implementations can provide “standard” GIS processing capabilities (e.g., [2]), or can deploy complex modelling tools [7] [14].

The Geo-Hydrological Hazard Protection Research Institute (IRPI), of the Italian National Research Council (CNR), has implemented an SDI dedicated to disseminating geospatial data and information on geo-hydrological hazards, in Italy. The SDI exploits Open Source Software [24] and in particular: (i) PostgreSQL 9.0 + PostGIS 1.5 (<http://www.postgresql.org/>, <http://postgis.refractor.net/>), (ii) GeoServer 2.0.2 (<http://geoserver.org>) (iii) GeoNetwork OS 2.6 (<http://geonetwork-opensource.org/>), (vi) ExtJS 3.4 + GeoExt 1.0 + OpenLayers 2.11 (<http://www.sencha.com/products/extjs>, <http://geoext.org/>, <http://openlayers.org/>). The system is implemented using more than 20 virtual machines hosted on four physical servers. Moreover, to deliver map services to different types of users, the CNR IRPI SDI exploit a specific thematic Web portal (<http://geomorphology.irpi.cnr.it/map-services>) that portal gives access to geographical and thematic data and information on landslides and floods, including: (i) landslide inventory

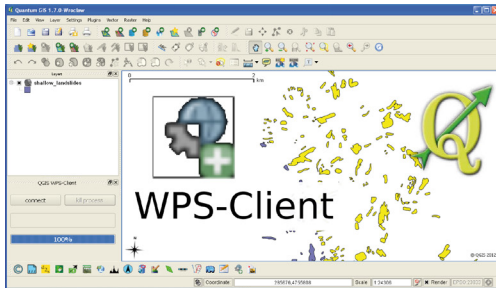
maps at different scales and covering different geographical areas, (ii) landslide susceptibility, hazard, and risk maps, (iii) information on historical landslide and flood events in Italy, and (iv) a prototype system to forecast the possible occurrence of rainfall-induced landslides in Italy [9] [21] [20].

SDIs provide chiefly data discovery, retrieval and visualization capabilities, i.e. cataloguing and data sharing services [6]. The published thematic maps are most commonly obtained through complex procedures implemented using different programming languages and software. The software is not always available to scientists, students, public administrators, or the general public. The adopted procedures are occasionally described in scientific articles, or technical reports, but rarely the software code is available. When the code is available, it may be difficult (or impossible) to use because of incompatibility/portability problems, or dependence on additional external libraries. Limited or incomplete understanding of software, procedures, or programming languages, further restricts widespread use of the resources.

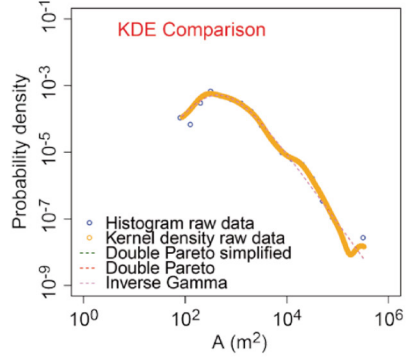
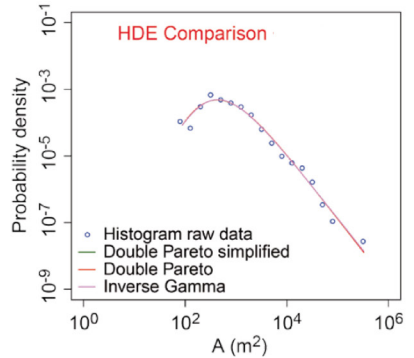
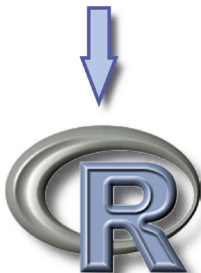
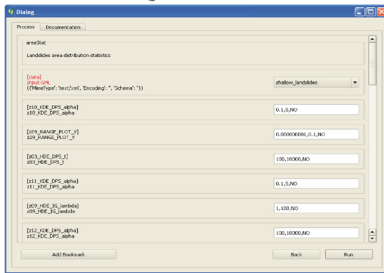
To overcome some of these problems, we have developed and deployed software tools, and associated processing procedures, to support geological and geomorphological mapping. Based on GRASS GIS 7 (grass.osgeo.org/grass70/), R (www.r-project.org/), and Python Web Processing Service (PyWPS), the tools are implemented as standard WPS, and allow for: (i) statistical modelling of the distribution of landslide areas, (ii) estimation of the attitude of bedding planes from the corresponding bedding traces, (iii) production of maps showing the geometrical relationship between bedding planes and terrain slopes, and the (iv) automatic delineation of hydrological slope units.

The first tool estimates the probability density and the frequency density of landslide areas shown on landslide inventory maps. The tool implements parametric and non-parametric statistical approaches, including Histogram Density Estimation (HDE), Kernel Density Estimation (KDE), and Maximum Likelihood Estimation (MLE), and produces a document listing statistics of the original data, parameters of the estimated distribution functions, and charts showing the estimated distribution functions [19] (see Figure 1).

FIGURE 1



↓
PyWPS



The second tool can be used to determine the attitude (dip direction and angle) of single or multiple bedding planes in an area, using a digital representation of the terrain (Digital Elevation Model, DEM) and a map showing “bedding traces” i.e., lines representing the geometrical intersection between bedding planes and terrain, obtained chiefly through the visual interpretation of aerial photographs, or through field mapping. The tool outputs a geographical vector layer of points and associated information on the dip direction, the dip angle, and the related uncertainty, for each point [16]. The output can be used to prepare structural maps.

The third tool implements a method for the interpolation of bedding planes to obtain spatially distributed information on the geometrical relationship between bedding planes and local terrain slope [22]. The tool outputs a geographical vector layer of polygons classified on five morpho-structural domains, including: (i) “anaclinal”, (ii) “orthoclinal”, (iii) “cataclinal under-dip”, (iv) “pure cataclinal”, and (v) “cataclinal over-dip” slopes. The output aids geological field mapping [1], and can be used as an input terrain variable to determine landslide susceptibility [3].

The last tool partitions a territory into slope-units, hydrological terrain subdivision bounded by drainage and divide lines [3], exploiting a digital representation of the terrain (DEM). The size and geometrical characteristics of the slope-units are controlled by user-defined modelling parameters, including the minimum half-basin surface area, and the maximum aspect standard deviation. The terrain subdivision obtained by the tool can be used to model landslide susceptibility [3] or hazard [8].

[1] BARCHI, M., GUZZETTI, F., FEDERICO, C. Reconstructing thrust geometry by map analysis techniques: three examples from the Central Apennines (Italy). *Bollettino Società Geologica Italiana* (1991) 110, 833-841.

[2] BÉJAR, R., LATRE, M.A., LOPEZ-PELLICER, F.J., NOGUERAS-ISO, J., ZARAZAGA-SORIA, F.J., MURO-MEDRANO, P. R. SDI-Based Business Processes: A Territorial Analysis Web Information System in Spain. *Computers & Geosciences* (2012). <http://dx.doi.org/10.1016/j.cageo.2012.04.010>.

- [3] CARRARA, A., CARDINALI, M., DETTI, R., GUZZETTI, F., PASQUI, V., REICHENBACH, P. GIS Techniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landform* (1991) 16, 427-445.
- [4] DE JESUS, J., WALKER, P., GRANT, M., GROOM, S. WPS orchestration using the Taverna workbench: The eScience approach. *Computers & Geosciences* (2011) doi:10.1016/j.cageo.2011.11.011.
- [5] FENOY, G., BOZON, N., RAGHAVAN, V. ZOO-Project: the open WPS platform. *Applied Geomatics* (2012) doi: <http://dx.doi.org/10.1007/s12518-011-0070-0>.
- [6] GIULIANI, G., NATIVI, S., LEHMANN, A., RAY, N. WPS mediation: An approach to process geospatial data on different computing backends. *Computers & Geosciences* (2011) doi:10.1016/j.cageo.2011.10.009.
- [7] GRANELL, C., DIAZ, L., GOULD, M. Service-oriented applications for environmental models: Reusable geospatial services. *Environmental Modelling & Software* (2010) 25, 182-198.
- [8] GUZZETTI, F., REICHENBACH, P., CARDINALI, M., GALLI, M., ARDIZZONE, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* (2005) 72, 272-299, doi:10.1016/j.geomorph.2005.06.002.
- [9] GUZZETTI, F., ROSSI, M., PERUCCACCI, S., BRUNETTI, M.T., MARCHESINI, I., ARDIZZONE, F., BALDUCCI, V., BIANCHI, C., CARDINALI, M., FIORUCCI, F., MONDINI, A., REICHENBACH, P., SALVATI, P., TONELLI, G., DELLO BUONO, D., IZZI, F., AMATO, L., LA SCALEIA, G., MAIO, D., PAGLIARA, P., DE BERNARDINIS, B. A national early warning system for rainfall-induced landslides in Italy. *American Geophysical Union, Fall Meeting 2010*, abstract #NH31C-02. 2010
- [10] HANSEN, H.S., SCHRØDER, L., HVINGEL, L.T., CHRISTIANSEN, J.S. Towards Spatially Enabled e-Governance: A Case Study on SDI implementation. *International Journal of Spatial Data Infrastructures Research* (2011), 6, 73-96.
- [11] JARAR, O.H., LOWNER, R., BENAABIDATE, L., WACHTER, J. HydrIS: An open source GIS decision support system for groundwater management (Morocco). *Geo-Spatial Information Science* (2009) 12(3), 212-216.
- [12] KÖHLER, P. WÄCHTER, J. Towards on Open Information Infrastructure for Disaster Research and Management: Data Management and Information Systems Inside DFNK. *Natural Hazards* (2006) 38,1-2, 141-157.

- [13] KRESSE, W., DANKO, D. M., FADAIE, K. STANDARDIZATION. In Kresse and Danko (Eds.): *Springer Handbook of Geographic Information*. Springer-Verlag Berlin Heidelberg. 2012
- [14] LECCA, G., PETITDIDIER, M., HLUCHY, L., IVANOVIC, M., KUSSUL, N., RAY, N., THIERON, V. Grid computing technology for hydrological applications. *Journal of Hydrology*, (2011) 403 (2011) 186-199.
- [15] MANSOURIAN, A., RAJABIFARD, A., VALADAN ZOEJ, M.J., WILLIAMSON, I.P. Using SDI and Web-Based System to Facilitate Disaster Management, *Computers & Geosciences* (2006) 32, 303-315, Canada.
- [16] MARCHESINI, I., SANTANGELO, M., FIORUCCI, F., CARDINALI, M., ROSSI, M., GUZZETTI, F. A GIS method for obtaining geologic bedding attitude. *Proceedings of the Second World Landslide Forum - Putting Science into Practice*, Rome, 3-9 October 2011. In press.
- [17] PERCIVALL G. Progress in OGC Web Services Interoperability Development. In Di, Liping & Ramapriyan H.K (Eds.): *Standard-Based Data and Information Systems for Earth Observation. Lecture Notes in Geoinformation and Cartography* (2010) 37-61. Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-540-88264-0_4.
- [18] RAUP, B., RACOVITEANU, A., KHALSA, S.J.S., HELM, C., ARMSTRONG, R., ARNAUD, Y. The GLIMS Geospatial Glacier Database: a New Tool for Studying Glacier Change. *Global and Planetary Change* (2007) 56, 101-110.
- [19] ROSSI, M., CARDINALI, M., FIORUCCI, F., MARCHESINI, I., MONDINI, A. C., SANTANGELO, M., GHOSH, S., RIGUER, D.E.L., LAHOUSSE, T., CHANG, K.T., GUZZETTI F. A tool for the estimation of the distribution of landslide area in R. *Geophysical Research Abstracts Vol. 14*, EGU2012-9438-1, 2012 EGU General Assembly 2012
- [20] ROSSI, M., PERUCCACCI, S., BRUNETTI, M.T., MARCHESINI, I., LUCIANI, S., ARDIZZONE, F., BALDUCCI, V., BIANCHI, C., CARDINALI, M., FIORUCCI, F., MONDINI, A.C., REICHENBACH, P., SALVATI, P., SANTANGELO, M., BARTOLINI, D., GARIANO, S.L., PALLADINO, M., VESSIA, G., VIERO, A., ANTRONICO, L., BORSELLI, L., DEGANUTTI, A.M., IOVINE, G., LUINO, F., PARISE, M., POLEMIO, M., GUZZETTI, F., TONELLI, G. SANF: a national warning system for rainfall-induced landslides in Italy. *Proceeding of the 11th International Symposium on Landslides (ISL) and the 2nd North American Symposium on Landslides (NASL)* 2012. Banff, Alberta, Canada, 2-8 June 2012. In press.

- [21] SALVATI, P., MARCHESINI, I., BALDUCCI, V., BIANCHI, C., GUZZETTI F. Management and publication of a continuously updated catalogue of geo-hydrological events with consequences to the population in Italy. *Proceedings of the Second World Landslide Forum - Putting Science into Practice*, Rome, 3-9 October 2011. In press.
- [22] SANTANGELO, M., BUCCI, F., CARDINALI, M., MARCHESINI, I., ROSSI, M., GUZZETTI F. Morpho-structural influences on landslide pattern and distribution: Grass GIS tool application. *Geophysical Research Abstracts Vol. 14*, EGU2012-12457, 2012 EGU General Assembly 2012.
- [23] SONG, X., RUI, X., HOU, W., TAN, H. An OGC standard-oriented architecture for distributed coal mine map services. *Journal of China University of Mining and Technology* (2008) 18(3) 381-385.
- [24] STEINIGER, S., HUNTER, A.J.S. Free and Open Source GIS Software for Building a Spatial Data Infrastructure. In: *Geospatial Free and Open Source Software in the 21st Century* (2012). Bocher E. & Neteler M. (Eds.), Part 5, 247-261.
- [25] STRAIN, L., RAJABIFARD, A., WILLIAMSON, I. P. Marine Administration and Spatial Data Infrastructure. *Marine Policy Journal* (2006) 30 (4), 431-441.
- [26] TUCHYNA, M. Establishment of spatial data infrastructure within the environmental sector in Slovak Republic. *Environmental Modelling & Software* (2006) 21, 1572-1578.
- [27] ZHAO, P., FOERSTER, T., YUE, P. The Geoprocessing Web. *Computers & Geosciences* (2012) <http://dx.doi.org/10.1016/j.cageo.2012.04.021>.

The Challenge of Geospatial Big Data Analysis

Authors

- Teerayut Horanont, University of Tokyo, *Japan*
- Apichon Witayangkurn, University of Tokyo, *Japan*
- Shibasaki Ryosuke, University of Tokyo, *Japan*

KEYWORDS : Pervasive geosensing, Spatial-temporal data, Big data analytics, Open source solutions, Cloud computing

Abstract - This paper discusses a new potential use of massive call phone location data for analyzing the urban system. Increasingly, large amount of spatial temporal information is becoming available with the help of technologies such as satellite-based positioning technologies, sensor networks and the penetration of the mobile phones. The mobile GPS data collected during year 2010-2011 and the special event of 311 Great Japan Earthquake is used to demonstrate the use case. These data are concerning as “Geospatial Big Data” where commercial or proprietary GIS software could not offer a suitable support for this particular large data set. We developed a prototype platform using all open source solutions as new approaches to handling, processing and analysis of the data. The results show that open source software and libraries are capable and are promising solution to cope with very large-scale geospatial data challenge.

1. Introduction

Communication network enables cities to gather more high-quality human mobility data in a timely fashion than ever before. Thinking in term of “scales of networks”, mobile communication provides us an ideal solution to create a huge urban fabric in which urban populations simply become part of

a network. These networks can be determined as telematic, physical, or even social interaction when people are all engaged. This research underscores the critical need and the opportunistic of these territories as well as the way to construct a new approach to manage of very large-scale spatial-temporal data by using open-source solutions.

In principle, telecommunication infrastructure may be seen as providing a service to people. This means people and infrastructure interact, making new interfaces and ways of representing the existence of any entity that can be seen by the network. The footprints from this interaction clearly become a new source to unambiguously identify people in the real world and that of create “space-time travel” data. Another reality is that this new spatial temporal data generated from telecommunication domain are rapidly increasing at a speed surpassing the capacities of ordinary computer’s storage and computing capacity. How to handle such a large-scale geospatial dataset? This is a critical issue especially in spatial analysis domain. The ability to gain speed in data processing, data mining and data analytical support is a big concern of today.

In this study, a new type of mobile-GPS data has been collected for 1-year period from August 1, 2010 to July 31, 2011. Approximately 9.2 billion of GPS records from 1.56 million registered users in Japan are input into our system for analysis. The mobile-GPS is a service provided by a leading mobile phone operator in Japan. Technically, the mobile-GPS enabled handset position is measured every 5 minutes, the information is then sent through the network to perform specific analysis and provide services for the registered users. This time series GPS data is completely anonymous and could not be identified the individual.

2. Reviews

Cloud computing platform is suggested for processing such large data in range of terabytes to petabytes with dynamically scalable and virtualized resources [1,2]. Hadoop is an open source large-scale distributed data processing that are mainly designed to work on commodity hardware [3]

meaning it does not require high performance server-type hardware. Hive is a data warehouse running on top of Hadoop to serve data analysis and data query by providing SQL-like language called HiveQL [4,5]. Hive allows users familiar with SQL language to easily understand and able to query data.

Our previous study, Witayangkurn et al [6] provided a comparative assessment of very large scale spatial data processing by comparing performance of three techniques specific for mobile dataset which are PostgreSQL with PostGIS, Java application with Java Topology Suite (JTS), and Cloud computing platform using Hadoop. In this implementation, the Hadoop platform with customized spatial function on Hive, an SQL-like language for Hadoop, give a drastically increase of performance boost from the traditional spatial database.

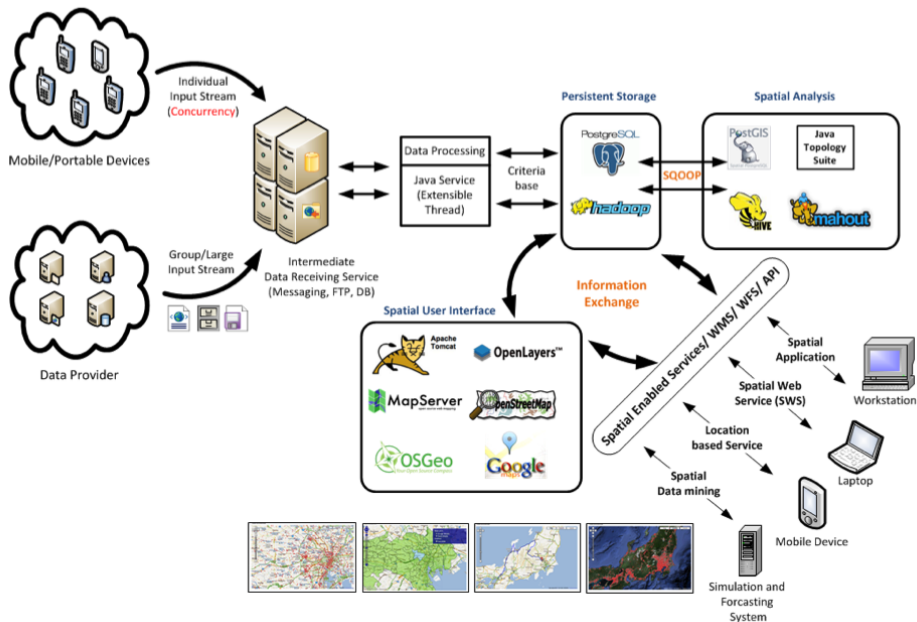
3. Implementation and System Overview

The GPS trajectory data of about 9.2 billion records and 600GB in size had been prepared to perform the analysis. In this research, the Hadoop Distributed File System (HDFS) and Hive is used to store and process the data. Generally, Hadoop/Hive did not support spatial query. However, Hive allows developers to create User-Defined Function (UDF) that could be any function based on the user requirement. We developed a custom function on top of Hive to manipulate and process the spatial temporal data. Figure 1 illustrates the overall system and basic idea in creating open source geospatial platform that do support the big data analysis and visualization. We have experience in conducting a system, which utilizes only spatial database (PostgreSQL/PostGIS) Horanont et al [7]. The previous system has several limitations and mainly from the constraint of high processing cost and calculation time regarding size of the data set. The proposed cloud computing technique dramatically increase the performance of spatial query [6] and therefore promising approach to address a new way of very large-scale geospatial processing.

4. Applications and Discussions

The March 11 Great Japan Earthquake event in 2011 was taken as our first analysis and visualization using the proposed platform. After a 9.0

FIGURE 1



magnitude earthquake struck the coast of Japan, 46 minutes later a massive tsunami flattened the Fukushima Daiichi nuclear power plant (DNPP). This has resulted in large-scale radiation leaks and eventually forced everyone living within 20 km to evacuate their homes. By the early time of evacuation period, it was not even known where most of the evacuees were. The local government in Fukushima said they didn't know where 40 percent of the residents around the Fukushima DNPP went [8]. We utilized this platform to demonstrate a near real-time monitoring of people movement in disaster areas since it is crucially important for developing an evacuation plan.

To demonstrate this scenario, the past 6 months of mobile-GPS data were used to calculate and find the numbers of people who live within the restricted area of Fukushima DNPP. The most often visited places during 0-6 am are determined as their home locations. By giving the same assumption, the data after March 11 were computed to estimate their evacuation places after the

nuclear accident. Theoretically, monitoring of the evacuation activity can be performed in near real time or in daily basis. Figure 2 (a, b, c, d) explain daily evacuation situation after March 11. Figure 3 shows the area where the evacuees moved during the first month. The line thickness connected between two cities defines the evacuation ratio. This evacuation ratio is an estimate of the fraction of the population exiting from a particular city within the restricted area to the city where they seek for shelter or temporary housing, which is defined in black dot on the map. Please note that more than 200 cities were detected as temporary places for the refugees who live within the restricted area although only the cities where evacuation ratio is greater than 5 percent are plotted on the map.

This initial study demonstrates the first use of mobile-GPS as pervasive emergency information retrieval tool and the results disclose part of human movement during the crisis. Managing and Mining of large-scale participatory data of mobile systems are crucial importance for the emergency management, particularly in the collection of real time data to provide accurate and complete picture of the current situation to the decision makers.

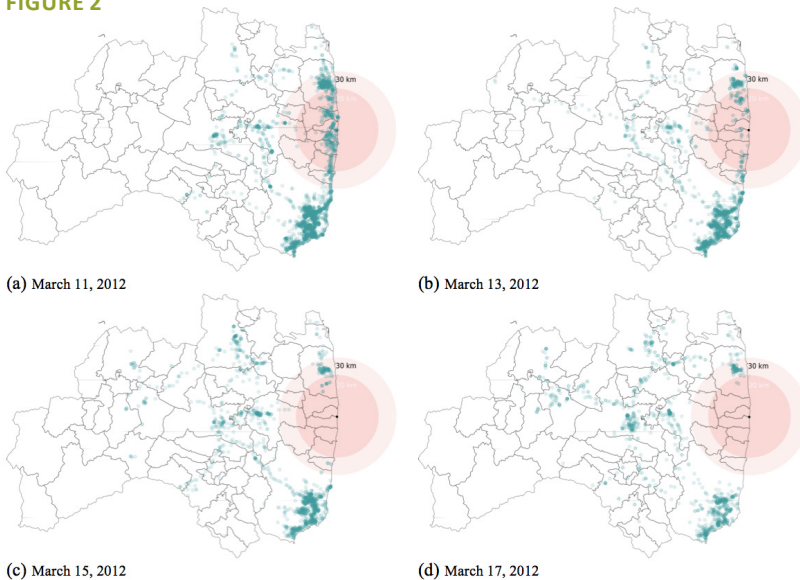
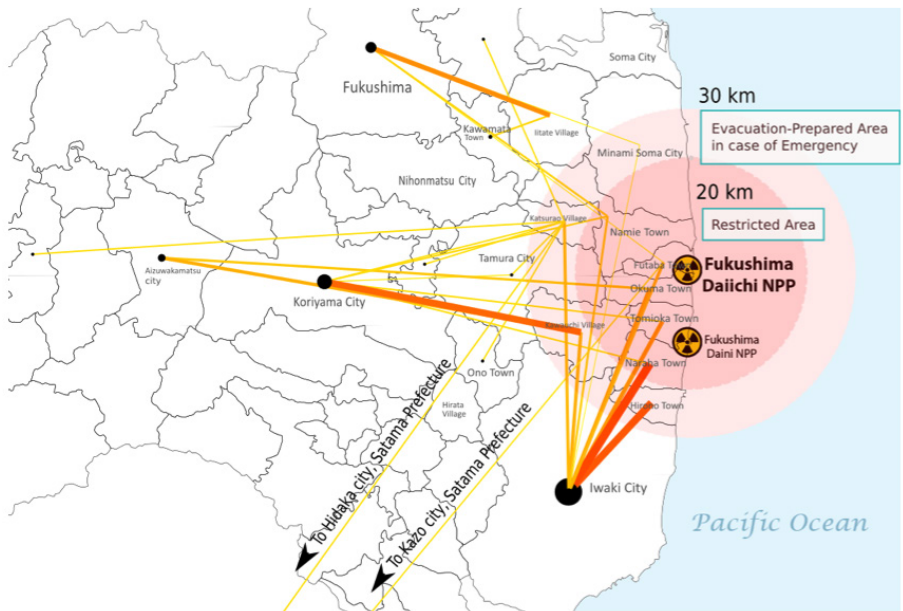
FIGURE 2

FIGURE 3



[1] YANG, J., AND WU, S. Studies on Application of Cloud Computing Techniques in GIS. In *IITA-GRS 2010: Proceeding of the IEEE Second IITA International Conference on Geoscience and Remote Sensing* (China, 2010), pp. 492-495.

[2] BUYYA, R., et al. 2008. Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Generation Computer Systems archive* 25 (2009), 599-616

[3] Hadoop Project. Retrieved March 11, 2012 from <http://hadoop.apache.org/>

[4] Hive Project. Retrieved March 11, 2012 from <http://hive.apache.org/>

[5] THUSOO, A., et al. Hive - A Petabyte Scale Data Warehouse using Hadoop. In *ICDE 2010: Proceedings of the 26th International Conference on Data Engineering* (California, CA, USA, 2010), IEEE Press, pp. 996-1005.

- [6] WITAYANGKURN, A., HORANONT, T., SHIBASAKI, R. Performance Comparison of Spatial Data Processing Techniques for Large Scale Mobile Phone Dataset. In *COM.Geo 2012: Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications* (New York, NY, USA, 2012), ACM Press.
- [7] HORANONT, T., AND SHIBASAKI, R. An Implementation of Mobile Sensing For Large-Scale Urban Monitoring. In *UrbanSense'08: Proceeding of the International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems* (North Carolina, NC, USA, 2008).
- [8] HAYS, J. 2012. Fukushima Evacuees. Retrieved March 11, 2012 from <http://factsanddetails.com/japan.php?itemid=2288&catid=26&subcatid=162>

Collaborative authoring and polypublication of cartographic content

Authors

- **Olivier Ertz**, HEIG-VD, University of Applied Sciences, *Western Switzerland*
- **Julien Le Glaunec**, University of Applied Sciences, *Western Switzerland*
- **Erwan Bocher**, IRSTV FR CNRS-2488, *France*

KEYWORDS : cartography, standards, collaborative, authoring, polypublication

Background and objectives

The double motivation of this project starts from the observation that there is an increasing number of producers and consumers of maps. Even, the term of «prosumer» has a growing acceptance to illustrate this phenomena [1] which is driven by the GeoWeb technologies.

Nonetheless, linked to this democratization of mapping techniques and usages, many experts address the question of credibility and quality of maps due to poor cartographic design [2]. This question initiates the first objective of this project which is about the experimentation of a collaborative authoring of cartographic content.

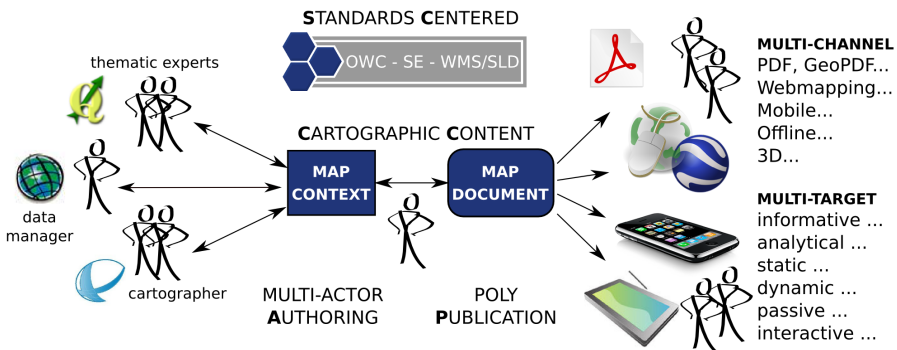
Moreover, map media for the masses is growing up. The use of maps within information content become a usual practice, again facilitated by GeoWeb technologies. This trend comes with a changing context concerning final user needs and diversity of diffusion channels. Thus, the second objective is

to experiment a multi-target and multi-channel platform for polypublication of cartographic content.

These two objectives, both dedicated to cartographic content and based on open standards, are illustrated below (Figure 1).

FIGURE 1

Standard Centered Authoring and Publication of Cartographic Content



Collaborative authoring

The Web 2.0 evolution contributes to bring to reality the initial purpose of the web pioneers to make it a collaborative space where people can interact [3]. In this regard, the purpose is to build a collaborative system for the authoring of cartographic content, so as to allow all actors of the creation process to avail its own skills at fair value (data manager, thematic expert, cartographer). The hypothesis is that a collaborative context is a way to reinforce the quality of maps thanks to the combining of skills within the teamwork.

The present distributed approach relies on the interoperability between cartographic authoring tools regardless of the used platform. Thus, each actor has the choice to keep her/his favorite tool. Thus, the actor in charge of cartography within the teamwork is then able to concentrate on the cartographic aspects so as to guarantee the quality of the cartographic message.

Standards for the interoperability of cartographic content is required to realize such an environment. Especially portrayal interoperability which concerns the possibility to share cartographic descriptions between tools while, of course, keeping exactly the same cartographic rendering. Since 2000, the Open Geospatial Consortium (OGC) has designed such a standardized symbology model which is known as the Symbology Encoding specification (SE 1.1) [4]. Such a specification has to be able to describe a large range of cartographic representations, but the current state is quite poor to give a complete answer to this. Therefore, the present project contributed to restart activities within the OGC Symbology Encoding Standard Working Group [5]. By leading the standardization work and combining several Change Requests, there is now a draft specification for a next release of the standard. The symbology model includes new possibilities for cartographic representations: compound strokes, composite graphics, symbol charts like pie charts and histograms, absolute units of measure etc. In parallel to the enrichment of the rendering capabilities, the project has contributed to push an experimental addition to the symbology model which is about semantic [6].

Finally, the second stage for a collaborative authoring environment concerns the sharing itself by the creation of a common repository of maps to share within the teamwork. Such a sharing point is relevant so as to put oil in the collaborative gears, including: a web API with the usual operations (creation, retrieving, update and deletion of a map context) and a comments system with automatic notification to support the interaction between the actors during the creation process, just like a blog or a forum would offer.

Polypublication of cartographic content

We call the result of the cartographic authoring a map context, which has a final destination defined by different user needs and different publication channels. This introduces the second purpose of the project: the polypublication of cartographic content on the base of a unique standardized map document. The hypothesis is that a publication platform natively designed for polypublication will help to rationalize and improve the publication.

Polypublication is first about multi-target publication which means a map document can be published in different ways according to the needs of the targeted user (informative for the reader, analytical for the decision-maker). The publication can vary and conform according to a layout (e.g. a corporate presentation template) or «functionalities» depending on the user skills. Secondly, polypublication is about multi-channel which means to adapt a map document to the diversified characteristics of the final medium of diffusion like a paper or screen size, a resolution and the level of interactivity. The multiplication of the users profiles and these medium characteristics explains why it might make sense to adopt a polypublication approach from the start of a cartographic communication project.

Given that publication work is a set of tasks reserved to an expert of the domain, a new actor has to be introduced within the teamwork, we call it the cartopublisher role which requires above all to be aware of publication constraints. Moderation of the cartographic content under authoring is the first task for this role. The actor interacts with the teamwork to check whether all information are described within a map context: a title, a relevant map legend, authors, etc. The second task is about enrichment of the map context with more multimedia content (more text, articles, photos, diagrams, hyperlinks, etc) to back up the understanding of the cartographic message. The cartopublisher uses its own publication tools to accomplish the preparation work for publication. As soon as done with these tasks, the map context can be considered as a map document ready for polypublication.

For the publication of the map document, this project experiments a new way to design a platform by taking into account the polypublication concepts detailed above. Indeed, current platforms are rarely designed with polypublication in mind from the start. Often these are composed of a set of lazy coupled and juxtaposed technologies trying to adapt the content on a case by case basis without a global design strategy. Therefore, the present approach is rather designed through a publication model composed of nested levels based on channel-independent components, we call them «target components».

The first level defines components in charge of cartographic «functionalities». The various possibilities of combining components allow to bring answers to various users profiles.

Examples of components:

- MAPSHORTCUT: it offers shortcuts to the user to navigate to predefined map views (parameters: one bounding box per predefined view)
- INFOLAYER: the user is able to get information about the features of a map layer (parameters: the feature properties and the title of each property to present)

The approach is then completed by a second level which is to decline the appropriate adaptation for each medium considering its native characteristics, taking advantages of ones and reducing drawbacks of others.

Examples of adaptations:

- for MAPSHORTCUT: given an interactive channel, the main map view is changed on each selection of a shortcut on the user interface; given paper medium, the publication is composed of several map view instances, one for each shortcut.
- for INFOLAYER: given an interactive channel, each click on the map displays the usual «bubble information» (parameters: search radius around the location of click or touch); given paper medium, a data table is presented within the document including a solution for the user to be able to match each line of the table with the right feature on the map, e.g. through the addition of identification labels on the map (parameters: the property to use as a label).

Finally, we understand that there are «global» target parameters and «specific» channel parameters, which is exactly a consequence of the nested approach (Figure 2).

FIGURE 2

A nesting and configurable approach of polypublication

```
name: Publication générique
description: Publication générique
modules:
```

```
  mapView:
    baseLayer: none
  layerSwitcher:
    layerPlan: multiple
    transparency: 1
```

«target components» to combine to create a publication type for a given final user type (includes the setting of global parameters)

```
channels:
```

```
  webMap:
    # in apps/frontend/templates
    template: webmapGenericTemplate
    modules:
      mapView:
        # in data/smarty/templates
        template: webMapGenericTemplate
        navigation: true
        navigationUi: false
      layerSwitcher:
        # in data/smarty/templates
        template: webMapGenericLayerSwitcherTemplate
        transparencyUi: true
```

Nesting: for each channel, the target component has an appropriate and parametrable adaptation

```
pdf:
```

```
  # in apps/frontend/templates
  template: pdfGenericTemplate
  modules:
    mapView:
      # in data/smarty/templates
      template: pdfGenericTemplate
    layerSwitcher:
      # in data/smarty/templates
      template: pdfGenericLayerSwitcherTemplate
      tableOfContentLinks: true
```

Nesting: for each channel, the target component has an appropriate and parametrable adaptation e.g. for PDF output, a table of content with hyperlinks

The aim of the project is not to reinvent the wheel and the technologies in link with publication of cartographic content. Instead, it is to adopt a strategy oriented from the start towards polypublication through the abstraction of a target level (target components).

Results and future

The strong focus on standards allows at one side the required interoperability to share cartographic content. At the other side, it allows to build a polypublication engine based on a unique and standardized source. Also, it helps to rationalize software development without technology dependency. The polypublication experimentations based on a nested publication design have been able to bring a proof of the concept with promising results. A prototype implementation has been started using the open source Symfony framework [7].

From the start, the team project has shown one's willingness to contribute to standardization activities. The aim was to reactivate an upcoming major topic at the OGC and to boost dissemination of the research work results. Given the innovation requirements to push forward the OGC SE specification, the development of a reference implementation has been initiated with OrbisGIS [8] to support these activities and to which other implementations can be evaluated. The openness of the open source OrbisGIS software is helpful to ensure maximum transparency of the specification.

Finally, collaborative authoring is a typical use case of OGC SE standard, but due to spatial data infrastructure rules and directives which are set up from local to global (LGéo in Switzerland [9], INSPIRE in Europe[10]), we state that standardization still requires more involvement of academic bodies to progress.

[1] FABY, H., KOCH, A. From maps to neo-cartography. Proceedings of the 3rd International Conference on Cartography & GIS (Nessebar, Bulgaria, 2010)

- [2] DAS, T., KRAAK, M. Does neogeography need designed maps? Proceedings of the 25th International Cartographic Conference (Paris, France, 2011)
- [3] LANINGHAM, S. 2006. developerWorks Interviews: Tim Berners-Lee. IBM developerWorks web site. Retrieved October 10, 2012 from <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html>
- [4] OpenGIS Symbology Encoding Implementation Specification. Open Geospatial Consortium web portal. Retrieved October 10, 2012 from <http://www.opengeospatial.org/standards/se>
- [5] ERTZ, O. 2009. SLD/SE 1.2 Standard Working Group Charter document. Retrieved October 10, 2012 from <http://www.opengeospatial.org/projects/groups/sldse1.2swg>
- [6] BOCHER, E., ERTZ, O., LAURENT M., PETIT G., RAPPO D., HEGRON G. Cartography and standards : from design to final user. Proceedings of the 25th International Cartographic Conference (Paris, France, 2011)
- [7] About Symfony Open-Source PHP Web Framework. Official Symfony web site. Retrieved October 10, 2012 from <http://www.orbisgis.org/about:presentation>
- [8] About OrbisGIS. Official OrbisGIS web site. Retrieved October 10, 2012 from <http://www.orbisgis.org/about:presentation>
- [9] Assemblée fédérale de la Confédération suisse. 2007. Loi sur la géoinformation, LGéo. Website of Switzerland's Federal Authorities. Retrieved October 10, 2012 from http://www.admin.ch/ch/f/rs/c510_62.html
- [10] IOC Task Force for Network Services. 2011. Technical Guidance for the implementation of INSPIRE View Services. INSPIRE web portal. Retrieved from http://inspire.jrc.ec.europa.eu/documents/Network_Services/TechnicalGuidance_ViewServices_v3.0.pdf

WORKSHOPS

Dynamic space-time visualisation: an introduction to i2maps

Author

Christian Kaiser, Institute of Geography, University of Lausanne, *Switzerland*

KEYWORDS : Geovisualisation, WebMapping, Spatio-temporal Data

The objective of this workshop is to provide an overview of i2maps, an open-source geocomputing environment. i2maps provides a flexible programming framework for knowledge discovery from spatio-temporal data and web-oriented visualisation, providing a means to «enable your data to speak for themselves». It consists of two libraries, one written in Javascript, and one written in Python. The Javascript library is used for building the interactive user-interface. The Python library consists of a server-based API for linking data sources and spatio-temporal analysis modules to the Javascript library.

Exploratory visual analysis can be an efficient way of preliminary investigation and hypothesis elaboration. Combining modern visual analysis with state-of-the-art Web technologies and real-time data streams increases the accessibility to rich datasets and allows domain experts to explore complex relationships in an easy visual way. However, extracting patterns, meaning and knowledge from large, spatio-temporal datasets is a challenging task. Consequently, such datasets are typically underused in many applications [1]. i2maps tries to solve the problem of bringing rich spatio-temporal data sets into the Web browser, by offering a flexible and easy-to-use structure and algorithms to handle both databases and data streams. i2maps tries also to tackle the problem of integrating spatial analysis tools for big data sets and data stream into the visualisation tools by offering powerful incremental algorithms.

The architecture of *i2maps* is designed to allow for dynamic interactive visualisation of spatio-temporal data. The Javascript front-end handles the interactivity and requests the data from the server. The Javascript library allows visualisation of dynamic raster and vector layers, and offers an interactive timeline. The map framework is built on top of OpenLayers. The Python back-end connects to various data sources and sends the data in JSON format to the Web browser. The *i2maps* Python library builds on top of other state-of-the-art libraries such as Numpy for matrix operations and numerical computations, or GEOS for various GIS operations. The Pico framework allows calling Python functions directly from Javascript making the development of complex Web applications easier. The Python library also handles transparent connections to PostGIS or SQLite databases and handles the data transforms between the data source and the Javascript front-end. It is also possible to integrate various other data sources, such as file-based storage, or data streams. *i2maps* also provides a raster cube offering flexible storage of temporal and potentially high-dimensional raster data in an efficient way. The Javascript front-end supports visualisation and exploration of these spatio-temporal surfaces through interactive user queries.

i2maps also provides powerful incremental algorithms based on machine learning methods. These algorithms are designed to be trained from both static data sources and data streams, and can also be updated in near real-time. Kernel Recursive Least Squares Regression (KRLS) offers a flexible framework for kernel regression and spatial interpolation [2, 3]. Scalable Local Regression (SLR) is a spatial regression algorithm similar to Geographically Weighted Regression (GWR) [4], but suitable for big data sets [5]. It can typically be used to discover and visualise spatial heterogeneity. *i2maps* has also an implementation of Projectron++ [6], a powerful classification algorithm similar to a Support Vector Machine (SVM) [7], but also working as an incremental algorithm and suitable for big data sets.

This workshop gives first an overview of the architecture of *i2maps*, and then works through the process of setting-up a complete *i2maps* project, involving setting up the data sources, linking the input data to a customised

spatial analysis method, and providing the results as an interactive map/timeline in a Web browser. The process of building an interactive i2maps application will be illustrated using the example of an interactive weather app. This example requires storage of spatio-temporal sensor measurements such as temperature or rain, display of the sensor data on a map and in a timeline, and computing a KRLS spatial interpolation surface based on the sensor data. Other examples will be discussed in order to illustrate the different features of i2maps. Among these examples, an application for visualising real-time Twitter messages will be shown. An example of a thematic map featuring temporal data will also be demonstrated.

After the workshop, each participant should have the necessary knowledge to start a simple project on their own, by using the appropriate documentation and code examples. A pre-configured Ubuntu-based virtual image will be provided with i2maps and discussed example applications already installed. This image should enable the participants to get started with i2maps without working through the steps of installing the required components. Installation instructions will still be provided for users deciding to develop applications with i2maps. Options for deploying i2maps applications are also shown during the workshop. Workshop documentation will also be available for download.

i2maps is an initiative of the National Centre for Geocomputation of the National University of Ireland Maynooth. It is currently actively developed by a community of international developers. More information on i2maps is available at <http://ncg.nuim.ie/i2maps> and <https://github.com/christiankaiser/i2maps>.

[1] D. GUO, J. CHEN, A.M. MACEACHREN AND K. LIAO, A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP), *IEEE Transactions on Visualization and Computer Graphics* 12(6), pp. 1461–1474, 2006.

[2] Y. ENGEL, S. MANNOR AND R. MEIR, The Kernel Recursive Least-Squares Algorithm, *IEEE Transactions on Signal Processing* 52(8), pp. 2275–2285, 2004.

- [3] C. KAISER AND A. POZDNOUKHOV, Enabling Real-Time City Sensing with Kernel Stream Oracles and MapReduce, *First Workshop on Pervasive Urban Applications (PURBA)*, 12–15 June, San Francisco, CA, 2011.
- [4] A.S. FOTHERINGHAM, C. BRUNDSON AND M. CHARLTON, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, 2002.
- [5] A. POZDNOUKHOV AND C. KAISER, Scalable Local Regression for Spatial Analytics, *19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 1–4 November, Chicago, IL, 2011.
- [6] F. ORABONA, J. KESHET AND B. CAPUTO, The Projectron: a Bounded Kernel-Based Perceptron, *ICML'08: Proceedings of the 25th International Conference on Machine Learning*, 2008.
- [7] V. VAPNIK, *The Nature of Statistical Learning Theory*, 2nd Edition, Springer, 1999.

Exploratory Spatial Data Analysis with PySAL

Author

Sergio J. Rey, GeoDa Center for Geospatial Analysis and Computation. School of Geographical Sciences and Urban Planning, Arizona State University, *USA*

Description

A unique feature of this tutorial is the use of Python based software tools for spatial data analysis. Python is an object oriented scripting language that is gaining rapid adoption in the computational sciences. To facilitate this adoption within the GIScience community, Sergio Rey and Luc Anselin have collaborated on the creation of PySAL: Python Library for Spatial Analysis. Since its initial release in July 2010, PySAL has been downloaded over 16,000 times. This tutorial will introduce participants to version 1.4 of PySAL and provides hands-on experience with the exploratory spatial data analysis (ESDA) components of the library.

Objectives

This tutorial will offer participants the following :

- installation of Python tools for scientific computing
- introduction to Python for spatial data processing
- introduction to PySAL for exploratory spatial data analysis

Outline

1. PySAL Origins, Use and Overview
2. Package Installation
3. Tools: IPython and IPython Notebook

4. Spatial data processing with PySAL
 - a. Reading/Writing Shapefiles
 - b. Other file types
5. Visualization
 - a. matplotlib
 - b. choropleth mapping
6. Spatial Weights
 - a. Weights Construction
 - b. Weights Manipulation
 - c. Weights Conversion
 - d. Spatial Lag
7. Spatial Autocorrelation
 - a. Global Autocorrelation
 - b. Local Autocorrelation
8. Spatial Dynamics
 - a. Classic Markov Chains
 - b. Spatial Markov Chains
 - c. LISA Markov

Audience

GIScientists, researchers and students interested in using PySAL for computational scripting in spatial analysis.

Prerequisites

This tutorial is geared towards individuals who have a basic understanding of exploratory spatial data analysis. Previous experience with Python programming is recommended.

TinyOWS, the high performance WFS-T server

Authors

- **Olivier Courtin**, Oslandia, *France*
- **Vincent Picavet**, Oslandia, *France*

KEYWORDS : WFS-T, High performance, OGC, OsGeo, INSPIRE

Modern GIS architecture make use of webservices to leverage web-based scalable applications. This lead to lightweight infrastructures, which are versatile and efficient. Moreover, OGC webservices standardisation enables those architectures to be fully interoperable. The INSPIRE directives supports OGC standards and creates a strong demands on softwares implementing those standards.

One of the need for GIS is to be able to access vector data, edit them and contribute modifications back to the data source. Of course this need to be available on top of an HTTP layer, to be able to build web-based or desktop applications on top. This use case is the one targeted by the WFS-T norm (Web Feature Service - Transactionnal). Accessing data through WFS-T helps improving the interoperability of the solution, and let us easily build applications on top of it, but it adds a layer on top of the data.

WFS-T is therefore a service for which performances are really important, as well standard compliancy, and software footprint, which should be the smallest possible. Presented in OGRS 2009 for the 0.9 release, the TinyOWS project found wider use as an official WFS-T Server for QGIS developers, and is now part of the MapServer Suite itself (since MapServer 6.2). TinyOWS [1] is OpenSource, (MIT licence), and part of the OSGeo

foundation, as it is now directly integrated under MapServer's umbrella. TinyOWS strictly implements WFS OGC standards [2] (versions 1.0.0 and 1.1.0), and passes all OGC CITE tests, including OGC CITE beta [3], which represents around 1000 unit tests. TinyOWS can be configured with a really simple and concise XML file which is edited manually, or directly with a MapServer MapFile. So you could use the very same config file to deploy both MapServer (as a WMS server for instance) and TinyOWS (as a WFS-T server). TinyOWS deeply relies on the database to perform stored data access and manipulations. As for now the PostGIS backend is the only one available, SpatiaLite and Oracle Spatial are planned for futures releases.

Current TinyOWS dev team gathers about 10 developers sending patches, enhancing features, or performing heavy tests on this application. The development infrastructure has now migrated to the Mapserver infrastructure, and the Mapserver and TinyOWS model of development have been adapted accordingly. This is a good example of how a free software project can grow and find a place in the ecosystem where it fits right, and be integrated to harness the power of complementarity.

TinyOWS is really fast, and designed to avoid CPU and memory waste. Some users who face really intensive WFS-T usages (>3000 simultaneous users) succeed to make it works on a single server architecture ! TinyOWS is therefore the perfect solution to save CO2 emissions and headaches (at least if you mind WFS-T and love webservice)...

The forthcoming TinyOWS 2.0 will bring a lot of new features, and the development will continue to TinyOWS mainly in three general directions :

→ INSPIRE compliancy

TinyOWS intends to implement the «Download Service» part of the directive [4]. The download service aims at providing a webservice to let a user download raw geospatial data. INSPIRE download service implies the implementation of at least a subset of the WFS 2.0 standards. Among the main aspects to be added, paging, stored procedure calls, language are among the features. Versionning is also a strong enhancement to what is currently provided by WFS-T services.

→ More performance improvement

There is still room for TinyOWS performance improvements, and a few ideas are already promising. One of the directions is to tighten the integration between the various components. TinyOWS could for example be plugged directly into Apache as a module, leading to more performance. Another direction would be to allow the gzipping/shipping part of TinyOWS to be run on a different server.

→ Support more output formats

Some new formats can be added to TinyOWS, on top of the currently supported ones (GML, GeoJSON). Shapefile, X3D, KML to name a few could be great to have. Customizable response XSD schema, and even other related specifications such as SOS-T (OGC Sensor Observation Service) are other interesting features that could be added too.

→ Multiple database backend support

Another area of improvement is the database backend. Only PostGIS is currently supported, but other backends could be interesting. The first step is probably to add an abstraction layer to the database backend, then add support for new backends. Good candidates include Oracle Spatial, and Spatialite. The latter would enable to have a very light WFT-T solution with embedded data. This would also mean that multi-connexion support would be added too, fixing a compatibility problem with mapserver which already supports multiple backends.

The workshop will allow each participant to successfully :

- Install TinyOWS from sources (including libs and apps dependencies)
- Discover the first steps with TinyOWS demo installation and using QGIS as a WFS client
- Use QGIS as a WFS-T client, and check that data are really modified on database side
- Understand TinyOWS XML Configuration file in depth

- Know a step by step debug procedure when something does not work
- Use Fast-CGI and Apache Web server module, configure and compare performances
- Use MapServer as a WMS service and TinyOWS as WFS-T one, on a same dataset through a single common MapFile
- Use TinyOWS XSD schema to allow enhancing and extening GML output, for instance with CityGML output.
- Use TinyOWS GeoJSON output and OpenLayers as a WFS-T client

This workshop is therefore a complete showcase of TinyOWS software to leverage read/write webservices on GIS data. Full systems interoperability is a tough target to reach, and versatile but still highly compliant and robust software are a key factor in good system architecture implementation. Once interoperability is reached, then performance is the next hard point to get right, and is also a key factor for a successful INSPIRE / OGC web spatial data infrastructure. TinyOWS provides interoperability and performance, while keeping the configuration and complexity of the software very low.

[1] TinyOWS project homepage. TinyOWS website. Retrieved Sept 19, 2012 from <http://www.tinyows.org>

[2] OGC Web Feature Standard Implementation Specifications. Open Geospatial Consortium website. Retrieved Sept 19, 2012 from <http://www.opengeospatial.org/standards/wfs>

[3] OGC CITE Beta tests. OGC Compliance Testing Program website. Retrieved Sept 19, 2012 from <http://cite.opengeospatial.org/betaTesting>

[4] INSPIRE NETWORK SERVICES DRAFTING TEAM, 2009. Draft Technical Guidance for INSPIRE Download Services. Retrieved Sept 19, 2012 from [http://inspire.jrc.ec.europa.eu/documents/Network_Services/INSPIRE%20Draft%20Technical%20Guidance%20Download%20\(Versoin%202.0\).pdf](http://inspire.jrc.ec.europa.eu/documents/Network_Services/INSPIRE%20Draft%20Technical%20Guidance%20Download%20(Versoin%202.0).pdf)

Exploit Pleiades PHR data with the ORFEO ToolBox library

Authors

- Manuel Grizonnet, CNES, *France*
- Julien Michel, CNES, *France*

KEYWORDS : remote sensing, ORFEO, Pleiades, Open Source, image processing, OTB, Open access, Open science

Launched in December 2011, the first satellite of the Pleiades system allows of very high resolution images acquisition. This system is made of two «small satellites» (mass of one ton) offering a spatial resolution at nadir of 0.7 meters and a field of view of 20 kilometers. Moreover Pleiades 1 and 2 will offer exceptional roll, pitch and yaw (slew) agility, enabling the system to maximize the number of acquisitions above a given area. The second satellite will be launched in November 2012. ORFEO, the Pleiades Accompaniment Program, was set up by CNES, the French Space Agency, to prepare, accompany and promote the use and the exploitation of the images acquired by this very high resolution optical sensor, especially in public sector. The objectives of this program are :

- to assess the thematic capability of the Pleiades system to produce the various services required by end-users for distinct domains (defense, risks, cartography, hydrology, forestry, agriculture...) develop efficient tools to facilitate image information extraction by end-users
- to develop efficient tools to facilitate image information extraction by end-users

The Methodological Part of the ORFEO accompaniment program aims at preparing the use and exploitation of Pleiades sub-metric images. This preparation includes capitalizing on image analysis RD results and know-how, and assisting the work of the thematic group and more widely of

the future users by providing them with algorithms, methods and easily available tools to visualize and process the images for their needs [1]. To achieve this, CNES decided in 2005 to develop and maintain, in the frame of the ORFEO accompaniment program, the Orfeo ToolBox (OTB), an open-source remote sensing image processing library.

The Orfeo ToolBox is written in C++ on top of ITK, a medical image processing library, and interfaces seamlessly with other open-source image processing software such as GDAL GDAL or OSSIM. Orfeo ToolBox is released under the CeCILL Open source license (equivalent to GPL) and is available on multiple platforms (Windows, Linux and Mac OS X). OTB comes with a modular architecture and natural scalability to images size and number of bands of most algorithms, thanks to native parallel and on the flow processing.

Orfeo ToolBox provides a wide range of functionalities and algorithms. In addition to the basic image access, Orfeo ToolBox provides standard remote sensing preprocessing like orthorectification, radiometric calibration or pan-sharpening. But the richness of the library lies in image processing: common processing tasks like thresholding, band algebra or Fourier and wavelets transforms, features extraction, segmentation, change detection and classification are some of the several tasks it can do. Advanced processing like Object Based Image Analysis are also available.

The most straightforward way of using the Orfeo ToolBox is to write c++ processing chain on top of it, while being guided by the extensive developer-oriented documentation. However, there are other lesser known means to use it, dedicated to non-developers.

The first one is to use the OTB Applications framework [3]. It is a set of applications plugins that can be accessed through command-line, standalone QT graphical user interface, higher-level coding languages such as Python for instance, and plugins for the QGIS software. This framework can be easily extended in two ways :

- first, one can very easily write new application plugins and access them
- and second, one can easily use the application plugin interface to integrate all the available plugins into his own software environment

The second and most end-user oriented mean to access OTB functionalities is to use Monteverdi, an integrated software for everyday life image manipulation and analysis task, which gives access to some of the most popular functions in OTB. Originally intended as a support for remote sensing training course and capacity building activities, Monteverdi has gained a lot of interest from the end-users community as a complete FOSS tool. The French Space Agency (CNES) in collaboration with the French Institute for Research and Development are helping to the development of the use of remote sensing data by getting access to images (like Spot or Pleiades) but also by providing tools to manipulate these data. Monteverdi was developed in 2009 in the frame of the first course in Antananarivo. We received a lot of interesting feedback after the course, requests for the implementation of new functionalities grew steadily and so the development of Monteverdi continued. The Monteverdi application consists in a smart architecture which allows building processing chains by selecting modules from a set of menus. It supports raster and vector data and integrates lots of OTB goodies. The architecture takes advantage of the streaming and multi-threading capabilities of the OTB pipeline. It also uses cool features as processing on demand and auto-magic file format I/O. The application is called Monteverdi [2], since this is the name of the Orfeo composer. This is also in remembering of the great (and once open source) Khoros/Cantata software.

Dealing with remote sensing images often imply several steps to be able to go from raw images to value added maps. Firstly with geometric and radiometric corrections of the data sources and after to use the appropriate algorithms. Moreover new capabilities and performances of the remote sensing systems like Pleiades imply new processing methods, or adaptation of existing methods in order to capture the essential information for the application. All these steps need efficient tools to build good processing work-flows in very close cooperation with the final users for better integration of new products in their systems.

This workshop based on all OTB tools will give a hands-on approach to perform general remote sensing image processing chain and analysis based on existing use cases using Pleiades data as input.

This tutorial will be divided in several technical sessions:

- Geometry : from sensor to cartographic projections with Pleiades data
- Radiometric corrections : from DN to TOA reflectance
- Feature extraction in VHR images
- Overview of segmentation methods in OTB
- Pixel based classification with Monteverdi
- Efficient Object Based Image Analysis of Pleiades data with the ORFEO ToolBox
- Elevation maps extraction from along-track Pleiades stereo pair images
- Access to OTB functionalities in Quantum GIS

Useful links

- ▶ <http://smc.cnes.fr/PLEIADES/index.htm>
- ▶ <http://www.orfeo-toolbox.org/otb/>
- ▶ <http://www.orfeo-toolbox.org/SoftwareGuide/>
- ▶ <http://www.orfeo-toolbox.org/CookBook/>
- ▶ <http://www.orfeo-toolbox.org/doxygen/>

Related publications

- ▶ Monteverdi Remote sensing software from educational to operational context - EARSEL 2010
- ▶ Open Source Remote Sensing: Increasing the Usability of Cutting-Edge Algorithms - GRSS Newsletter (Open Source Remote Sensing: Increasing the Usability of Cutting-Edge Algorithms (Emmanuel Christophe, Member, IEEE and Jordi Inglada, Member, IEEE)
- ▶ The Orfeo ToolBox : on the way to massive remote sensing - Sentinel-2



Object Labeling Application

Map Overview Full Resolution

Map Overview

Zoom

Pixel Description

Object Description

Objects Features Learning

Objects

- Roofs (18 objects)
- Water (0 objects)
- Vegetation (17 objects)
- Dark roofs (4 objects)
- Roofs (17 objects)
- Shadows (0 objects)
- Stairs (0 objects)
- Pools (0 objects)

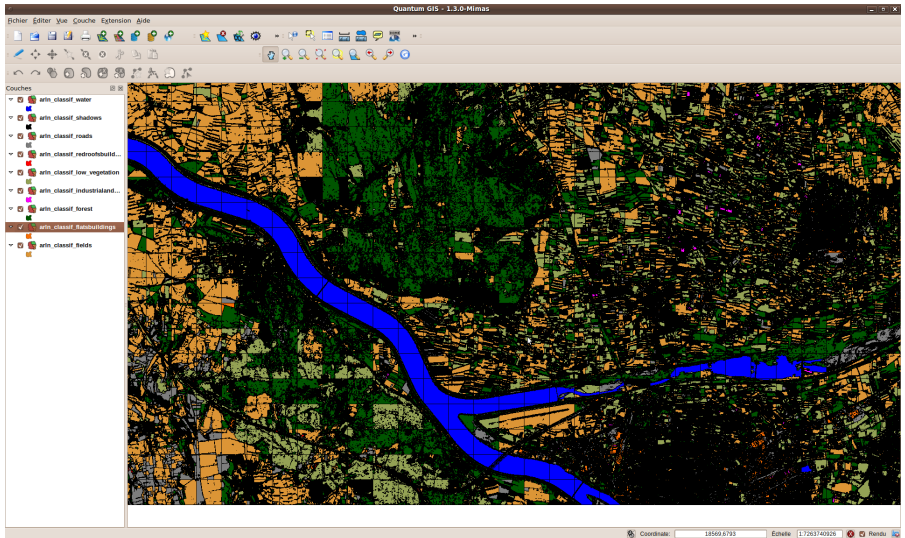
Label Color

New (Start with)

Add Remove Clear

objects	margin
851	100
46	910
446	616
2664	209
3817	174
2001	389
2228	281
1127	140
306	1400
47	200
1081	2670
1208	610
1100	1032
1708	1537
1107	42
884	2796
1492	382
382	1792
2084	2084
240	240
1820	1820
2097	2097
282	282
318	318
1087	42
42	1171
42	42
1029	1029
528	528
1792	1792
382	382

Remove Clear Change Clear



Illustrations

- Large scale segmentation using OTB and geospatial database in Quantum GIS
- Change detection in OTB using MAD/MAF algorithm
- Object labeling using object oriented SVM classification and active learning

[1] CHRISTOPHE, E., INGLADA, J. Open Source Remote Sensing: Increasing the Usability of Cutting-Edge Algorithms. *IEEE GRSS Newsletter* pp. 23-26, 2009.

[2] GRIZONNET, M., INGLADA, J. Monteverdi Remote sensing software from educational to operational context. *EARSEL Symposium*, 2010.

[3] MICHEL, J., GRIZONNET M. The Orfeo ToolBox : on the way to massive remote sensing. *Sentinel-2 Preparatory Symposium*, 2012.

Introduction to PostGis : data management and geoprocessing

Author

Andrea De Bono, UNEP GRID-Geneva, *Switzerland*

PostGIS is an open source, freely available, and fairly OGC compliant spatial database add-on for the PostgreSQL relational database server. It adds spatial functions such as distance, area, union, intersection, and specialty geometry data types to the database, allowing advanced spatial processing and querying.

Objectives

introduce the participant in the world of geodatabases by using a very powerful tool: PostGIS. The workshop will furnish to the participant basic notions allowing:

- create a geodatabase
- Load geodata and work with geometry and attributes
- Querying data and examples of geoprocessing
- View results in QGIS

Program

1. Introduction to PostGis
 - ▶ What is a Spatial Database ?
 - ▶ Spatial Column Types
 - ▶ Spatial Indexes and Bounding Boxes
 - ▶ Spatial Functions
 - ▶ What is PostGIS ?

- ▶ A brief history of PostGIS – Successive versions
 - ▶ Who uses PostGIS ?
 - ▶ What applications support PostGIS ?
2. A few words about the installation
 3. Creating a Spatial Database (using template /by command line)
 4. Loading spatial data
 - ▶ Import export (shp2pgsql, load a sql)
 - ▶ Spatially enable an existing dataset
 5. View data using QGIS, (ArcMap)
 - ▶ Add layers, style editor.
 6. Examples of simple queries
 7. Dumping data and databases
 8. PostGis functions overview

Audience

Beginner

Prerequisites

Basic GIS notions

Language

English or French

OrbisGIS : From GIS to connected geoservices

Authors

- **Antoine Gourlay**, IRSTV FR CNRS 2488, *France*
- **Alexis Guéganno**, IRSTV FR CNRS 2488, *France*
- **Gwendall Petit**, IRSTV FR CNRS 2488, *France*
- **Erwan Bocher**, IRSTV FR CNRS 2488, *France*
- **Olivier Ertz**, HEIG-VD, *Switzerland*
- **Julien Le Glaunec**, HEIG-VD, *Switzerland*

KEYWORDS : opensource, GIS, ogc, spatial analysis, cartography, SDI

OrbisGIS (<http://www.orbisgis.org>) is a geographical information system dedicated to scientific modelling and experimenting. OrbisGIS has been developed at the Institute for Research in Urban Science and Techniques (IRSTV FR CNRS-2488) (<http://www.irstv.fr>) since April 2007, within the Atelier SIG framework whose goal is to provide methods and tools to grasp the challenges of urban environments.

Distributed under a GPL 3 licence, OrbisGIS is a federating tool within a Spatial Data Infrastructure (SDI), gathering amongst the research units of the IRSTV all the methods and processed data linked to geographical information, irrespective of the research field they come from (sociology, civil engineering, urban architecture, geography, economy, environment...). Based on OGC standards, OrbisGIS tries to reduce the gap between GIS software and SDIs:

- on the one hand by making data consumption easy, using geospatial web services

- on the other hand by sharing data, thematic maps and processing chains

This workshop will be organised in 3 sections.

Part 1 : Discover the platform

This section will introduce some basic concepts and uses of the GIS platform. User will learn how to:

- install OrbisGIS and where to find documentation and help
- use OrbisGIS's UI: create a workspace and customize the user interface
- load flat files and DBMS tables
- connect to a set of WMS services
- use navigation tools and information tools
- query data, display attributes in or out of spatial context
- edit, create a spatial data source
- deal with styles and create simple thematic maps

Part 2 : From collaborative authoring to polypublication of cartographic content

At first, this part of the workshop aims at showing how a collaborative authoring environment can run thanks to common standards like OGC Map Context and OGC Symbology Encoding specifications. Attendees will experiment a multi-actor for authoring a common cartographic project through a shared remote repository.

Secondly, a new user role for the authoring process will be introduced : the «cartopublisher». Attendees will experiment a dedicated user interface to control and moderate through validation rules the cartographic content authored by the several actors. Also, attendees will experiment how the cartopublisher is able to enrich the content with supplemental items like text, images, diagrams, ... so as to stress the cartographic message.

Finally, attendees will discover a platform dedicated to the polypublication of cartographic content. Polypublication stands for multi-targets/

multi-channels publication. Multi-target concerns the configuration of the publication so as to fill the need of a final user according to a user profile and user skills. Multi-channel concerns the adaptation to the final channel of visualisation (webmapping, mobile, pdf, geopdf, ...). Attendees will experiment an integrated approach to configure such a multi-target/multi-channel publication engine.

Part 3 : From Spatial SQL to WPS

One of the main interests of OrbisGIS is its spatial SQL engine that allows building both simple and advanced spatial analysis. In this section we will show how to transform an SQL script and expose it as a WPS.

First of all, we will present the enhancements that have been made to this language. These improvements over the original Spatial SQL are the result of research funded by the french regional program GeoPAL (<http://www.geopal.org>) and ANR VegDUD. They cover performance improvements, mixed spatial-raster processing, improved syntax, and more.

Secondly we will introduce a simple meta-language to describe an SQL Script, and a way to publish the script as a WPS process directly from OrbisGIS. WPS provides a way to expose on-demand spatial processes as web services through a standard interface defined by the OGC. Attendees will experiment with the meta-language and the publishing interface with several kinds of script, from simple processes to advanced analysis.

Finally, attendees will discover how to seamlessly consume these WPS processes from OrbisGIS with various input and output types.

Introduction to QGIS and GRASS

Authors

- **Gregory Giuliani**, University of Geneva - enviroSPACE laboratory, *Switzerland*
- **Yaniss Guigoz**, University of Geneva - enviroSPACE laboratory, *Switzerland*

Geographic Information Systems (GIS) software have become extremely popular since global mapping services became available through the Web and mobile phones.

GIS acquire and process spatial data for a wide range of applications (e.g., natural resources management, archaeology, urban planning, environmental sciences, global change modelling, and cartography). Geospatial data is turned into geographic information through geoprocessing and visualisation. All GIS processes can now be conducted using open-source applications on Linux, Mac, or Windows-based platforms.

Indeed, proprietary packages are dominating the GIS market but convincing open-source software and data solutions are clearly emerging. These are especially attractive for students, GIS professionals, small and medium enterprises, companies and institutions in emerging countries, and international organizations.

This workshop aims at presenting basic functionalities of two majors Open Source Desktop GIS software: QGIS and GRASS. Through hands-on exercises participants will acquire the basic knowledge to be able to work with these software for handling vector and raster data (visualisation, processing), create maps, and documentation.

A CD with software, data, and relevant documentation will be provided to participants.

Programme

1. Short introduction on QGIS and GRASS
2. Discovering the QGIS community: users, developers, documentation, etc
3. Exercise 1: *Adding data with QGIS*
 - ▶ Learn how to add shapefile, geotiff, XY, GPS, WMS, and WFS data
4. Exercise 2: *Managing projections with QGIS*
 - ▶ Learn how to handle projections in QGIS and how to reproject data from Spain and China
5. Exercise 3: *Mapping with QGIS*
 - ▶ Discover the functionalities to prepare a map and export it in PDF and as an image
6. Exercise 4: *Data Analysis with QGIS*
 - ▶ Generate a population raster of a Jamaican region and extract basic statistics
7. Exercise 5: *Spatial queries with QGIS*
 - ▶ Compute selection by locations based on rivers and population data sets
8. Exercise 6: *Basic functionalities of GRASS and the QGIS-GRASS plugin*
 - ▶ Discover how to access GRASS GIS functionalities within QGIS

Quantum GIS as a platform

Author

Vincent Picavet, *Oslandia, France*

At the age of web and mobile GIS application, it could be considered a strange idea to focus on desktop GIS. But still, desktop software have a future and are still really pertinent when it comes to advanced analysis or visualization features, where web applications are currently limited.

Quantum GIS (QGIS) [1] is a user friendly Open Source desktop Geographic Information System (GIS) licensed under the GNU General Public License. QGIS is an official project of the Open Source Geospatial Foundation (OSGeo). It runs on Linux, Unix, Mac OSX, Windows and Android and supports numerous vector, raster, and database formats and functionalities.

The Quantum GIS project was officially born in May of 2002 when coding began. The idea was conceived in February 2002 when Gary Sherman began looking for a GIS viewer for Linux that was fast and supported a wide range of data stores. That, coupled with an interest in coding a GIS application led to the creation of the project. The first code was checked into CVS on SourceForge on Saturday July 6, 2002, and the first, mostly non-functioning release came on July 19, 2002. The first release supported only PostGIS layers.

Nowadays, QGIS has evolved a lot, and is gaining more and more attention from the geospatial community. From a simple PostGIS data viewer, it became a full-featured desktop GIS, with little lacking compared to the proprietary industry leaders.

Current version 1.8 and the next major 2.0 releases, are slowly changing the way QGIS is considered and used. It mutated from a desktop GIS to a GIS platform.

Wikipedia, the collaborative encyclopedia, defines a platform as «a term for technology that enables the creation of products and processes that support present or future development. It establishes the long-term capabilities of research & development institutes. It can be defined as a structural or technological form from which various products can emerge without the expense of a new process/technology introduction.» Latest QGIS versions correspond to this definition.

First of all, on a structural point of view, the opensource side of this software makes it ease collaboration and innovation. QGIS is GPL-licenced, and has the particularity of being completely community-driven. There are a lot of various contributors to the core of Quantum GIS, and decisions are taken after general community consensus. The Project Steering Committee's interventions are limited to the bare minimum. QGIS core source code is hosted on Github, available to fork for anyone, and all development versions are hackable. The QGIS infrastructure is also open, and bug reports, advices, feature requests and patches can be sent seamlessly.

This openness is a strong point in making QGIS a hub for collaboration on GIS software, GIS processing and visualization, be it on ideas, code, or documentation. This freedom in the way QGIS is developed helps innovation to appear. Being part of OSGeo and Google Summer of code is another aspect of this openness in collaboration.

On the technical side, there are a lot of aspects that make QGIS a platform more than a simple software : it has an API, it is a GIS connector to many other components, it is extendable, it is open, it propose a rapid development model, and offers Python bindings.

Quantum GIS is developed in C++ on top of the QT graphical framework. QGIS has an API which is documented, and allows anyone to develop new components to enhance it, or even build a full C++ GIS application using only QGIS library.

QGIS acts as a connector to many other GIS components. It opens a lot of GIS formats, including databases, thanks to the GDAL/OGR library at its heart. Geometry processing is done with the GEOS library. A powerful connector to the GRASS GIS is also integrated into the software, letting the user benefit from all GRASS features and algorithms. A 3D globe is also available, based on OpenSceneGraph and OSG Earth libraries. Of course it is a client for various OGC standards and can connect to WMS, WMSC, WFS or WFS-T servers and more.

Quantum GIS is extendable in many ways. Through code source contribution for new core features obviously, but mainly through the plugin mechanism. QGIS allows a user to extend it through plugins, and particularly plugins written in Python. QGIS expose its API as Python bindings (known as PyQGIS), which let a user write powerful new custom features. Writing Python plugins to QGIS is easy and fast, therefore leads to great innovation. New plugins are added to the main repository every week, for classic processing as well as very unique features. Python is great as a glue language, and QGIS benefits from this characteristic, becoming more and more a framework on top of which new tools and new usages are developed.

The next step will boost QGIS into the platform world : Sextante integration will make the difference. Sextante in QGIS is the future QGIS geoprocessing framework. Ported from GvSig/Java to QGIS/Python, the framework design is already mature, and the code is already usable. Its aim is to provide a framework to develop new geoprocessing algorithm, while keeping the integration efforts very low. As soon as a first version has been out, we have seen a lot of module developers integrating their favorite algorithms into Sextante very easily. There are now more than 10 modules available for



QGIS LOGO

the framework, each of which provide dozens to hundreds of new geoprocessing algorithm. The main advantages of using the framework are automatic interface generation for options, seamless integration between all algorithms, even from different backends, automatic integration into the workflow modeler, and ease of development.

This new geoprocessing framework is still in beta, but will be intensively worked on in the next months, and there is no doubt that it will be one of the main highlights for QGIS 2.0.

These new features will enable anybody to contribute to QGIS, with a very easy way of integrating external contributions to QGIS geoprocessing framework. This, and the opensource aspect, will lead to more contributions, more links between communities, academics, researchers, private companies, public sector, developer associations, opendata addicts... In the GIS world and all custom domains needing geo-located features and data.

Opensource licence, community-driven project management, open technical framework, acting as a concentration point for every other related component makes QGIS not only a desktop GIS, but also a platform. And it is growing big in that direction.

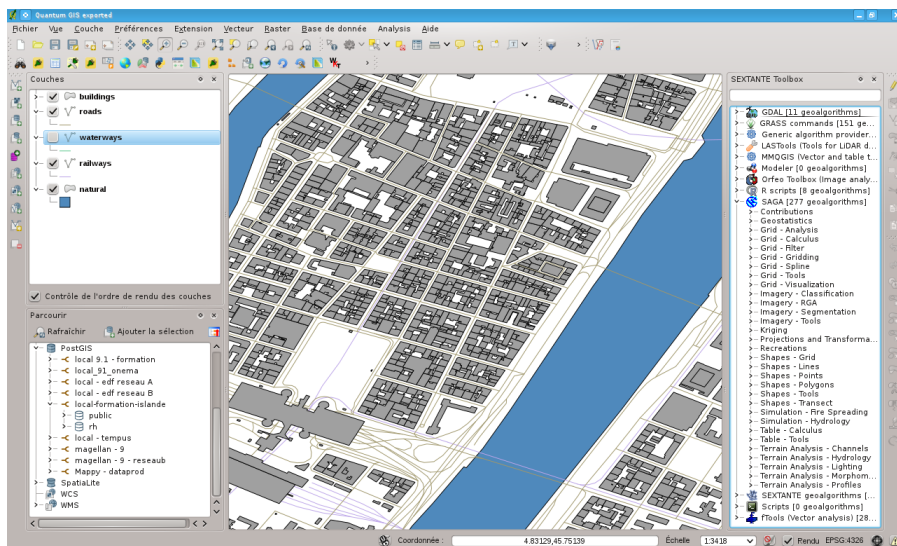


FIGURE 2

Quantum GIS with Sextante toolbox

[1] QUANTUM GIS DEVELOPMENT TEAM. Quantum GIS Project homepage. QGIS website. Retrieved August 31, 2012 from <http://qgis.osgeo.org>

DISCUSSION GROUPS

Bringing together international experts from academia, industry, government, and beyond, the OGRS symposium provides a neutral forum for participants to exchange point of views on the role of open source in research and education.

The symposium offers a session for discussion groups which are organized to address specific subjects in an open and comfortable context of sharing. Indeed, opportunities, challenges but also threats are particularly important for the significant topical subjects which are proposed:

→ *Academic reseach and open source software to serve open standards development*
(Chair: Prof Olivier Ertz)

The standardisation challenges for the geospatial community are manifold and academic research is continuously ongoing in these areas: research is leading to standards, and vice versa, standards are generating further need of research. Nonetheless, the white paper «Modernising ICT Standardisation in the EU - The Way Forward» (Commission of the European Communities, Brussels, 3.7.2009) highlights a major goal: «Fostering synergy between ICT research, innovation and standardisation». The definition of such a goal by the Commission clearly stresses the starting point of this discussion: the observation that in the geospatial domain, it seems there is a gap between research and standardisation, a gap in term of involvement of research bodies in standardisation processes. Thus, the motivation for this discussion is to argue and evaluate the reality of this observation and to identify opportunities to make a better connection between research and standardisation in the geospatial domain.

The outcomes of the discussion shall contain some arguments for all bodies, standards organizations, organizations funding research and innovation, researchers and research organizations and evaluation agencies. To feed the discussion, some questions might be considered, they include: Are all bodies aware of the economical benefits of standards and standardisation regarding research and innovation? While open source helps for the dissemination and transfert of knowledge and technologies through open innovation, what about open standards? Does research teams include some driving activities at the

heart of a standardisation body in their research planning? How do funding organizations and evaluation agencies consider standardisation activities in academic research? How do standardisation bodies consider academic research and do they favor the inclusion of research teams and their results? How does open source software support both research and standardisation?

- *What licence should be used to release software or data produced in an academic context? (Chair: Eric Grosso)*

Recently there has been massive public interest in the media's reports of an 'academic spring' in the UK, in other words the potential for an 'open access' to public funded research papers. As most research projects are financed by public funds that same public started to ask the question: «As a tax payer, can I obtain free access to the results of public research, and furthermore do I have the right to use the produced software or data resulting from this research?».

This subject therefore raises further questions for the academic population: «When research is funded with a public grant, what licence should be used to release software or data produced in an academic context?». With regard to the fact that academic researchers publish their work through articles, why not also publish the resulting software developed or data produced in this same context rather than by the usual practice of using a patent to release it?

In order to answer these questions, a discussion will be raised around the advantages and the drawbacks of freely sharing software and data.

- *Spatial Data Infrastructures in developing countries: Is Open Source a solution? (Chair: Dr Gregory Giuliani, Yaniss Guigoz)*

Open source technologies have recently reached a good level of maturity and could help to address the current weaknesses or threats regarding SDI implementation in Africa. The goal of this discussion group is to propose ways for successful SDI implementations for environmental management in Africa in light of these recent advances of open source philosophy and technologies. We will consider in this discussion the

current situation in Africa in terms of the various components of the SDI framework: the status of the legislation on data sharing and SDI, the existence of executive bodies linked to SDI, the trend in teaching about SDI and more generally about GIS, and the current participation of African countries to international SDI initiatives such as GEOSS and Eye On Earth. The existing status of the technical enablers of SDI, such as the availability of broadband access, the variety of mobile devices to access data, and the potential for existing at a point to fully adopt international standards for data sharing, will also be taken into account.

→ *Urban remote sensing and environmental indicators (Chair: Prof Elizabeth Wentz)*

The goal of this discussion group is to introduce and discuss the use of remote sensing technology for monitoring and modeling cities. The group will engage in an extensive discussion around five themes: the nature of cities, the problems to be solved, remote sensing solutions, spatial data sources, and classification, and analysis tools. The focus will emphasize open source software tools especially during the discussion of classification and analysis tools.

Cities are large and permanent settlements of human populations reflecting the physical, cultural, economic, and social conditions of the people living there. The city form is a critical research area because more people are moving to and living in cities compared with past generations. What is unknown is the impact cities have on the global environment and how those impacts will continue to change. What is known is that as cities grow and change through urban densification and sprawl there is 'consumption' of land for urban purposes. These land transformations have a profound effect on the biophysical processes locally –urban climate, soil properties, water patterns. These effects scale up to regional and eventually global changes. Remote sensing can be a tool to monitor and model these multi-scale processes.

This discussion group was motivated by a workshop organized by the discussion leader on urban remote sensing held in Arizona, USA in April 2011. One of the primary outcomes of the workshop was to identify the gaps in knowledge and research opportunities in urban remote sensing. A second outcome was insight that remote sensing scientists needed to engage with a larger research community, such as the participants at this conference.

→ *PGIS, from crowdsourcing to decision-making : research challenges, perspectives and emergences (Chair: Prof Stéphane Roche, Dr Matthieu Noucher)*

Past decades were mainly characterized by a very top down and institution-driven production of authoritative Geographic information. With the development of geo-crowdsourcing, volunteered geographic information (VGI) and open data, more bottom up forms of production and diffusion of GI are gradually emerging. Yet the use of such user generated geo-contents as a support for decision-making processes is still a complex issue. Even if local policies are potentially strongly transformed by the development of the GeoWeb 2.0, it is not clear to what extent individual expressions (VGI are in a sense mainly individual) could produce collective representations. One of the main challenges for geomatics research is then to invent new innovative models to integrate geo-crowdsourcing with formal decision-making processes.

Operating mode

Each session of discussion are organised in four main steps, based on a simplified SWOT analysis (Strength – Weakness – Opportunity – Threat).

- **Step 1 (5 min.)** : Chair's brief introduction of the topic : Why the topic has been proposed ? What are in his own point of view the main stakes to be considered ?
- **Step 2 (5 min.)** : Each participants writes on a post-it 2 or 3 different keywords classified as either a : Strength / Weakness / Opportunity / or Threat. (These categories will be characterised with specific colored post-it).

- Step 3 (30 min.) : Each participant sticks his keywords post-it on a common keyboard and comments briefly his choice.
- Step 4 (20 min.) : General Discussion in order to synthesis results of the SWOT analysis and to identify the main challenges and some recommendations.

Outcomes from each discussion group are presented during the closing plenary session. They will be published online after the symposium.

Stay tuned on : <http://ogrs2012.heig-vd.ch/discussions>



Symposium proceedings

Open Source Geospatial Research & Education Symposium

This book contains the extended abstracts of all contributions presented during the Open Source Geospatial Research and Education Symposium (OGRS 2012) held in Yverdon-les-Bains, Switzerland, October 24-26, 2012. OGRS is a forum for the exchange of knowledge, new solutions, methods, practices, ideas and trends in the domain of geospatial information through the development and use of geospatial free and open source software (GFOSS) in both research and education. The open source approach has many positive impacts; this is what OGRS wants to demonstrate in the context of geospatial research and education. In this regard, these proceedings cover numerous research and educational topics, with education constituting a new field of interest for OGRS.

Edited by Olivier Ertz, Stéphane Joost & Marj Tonini