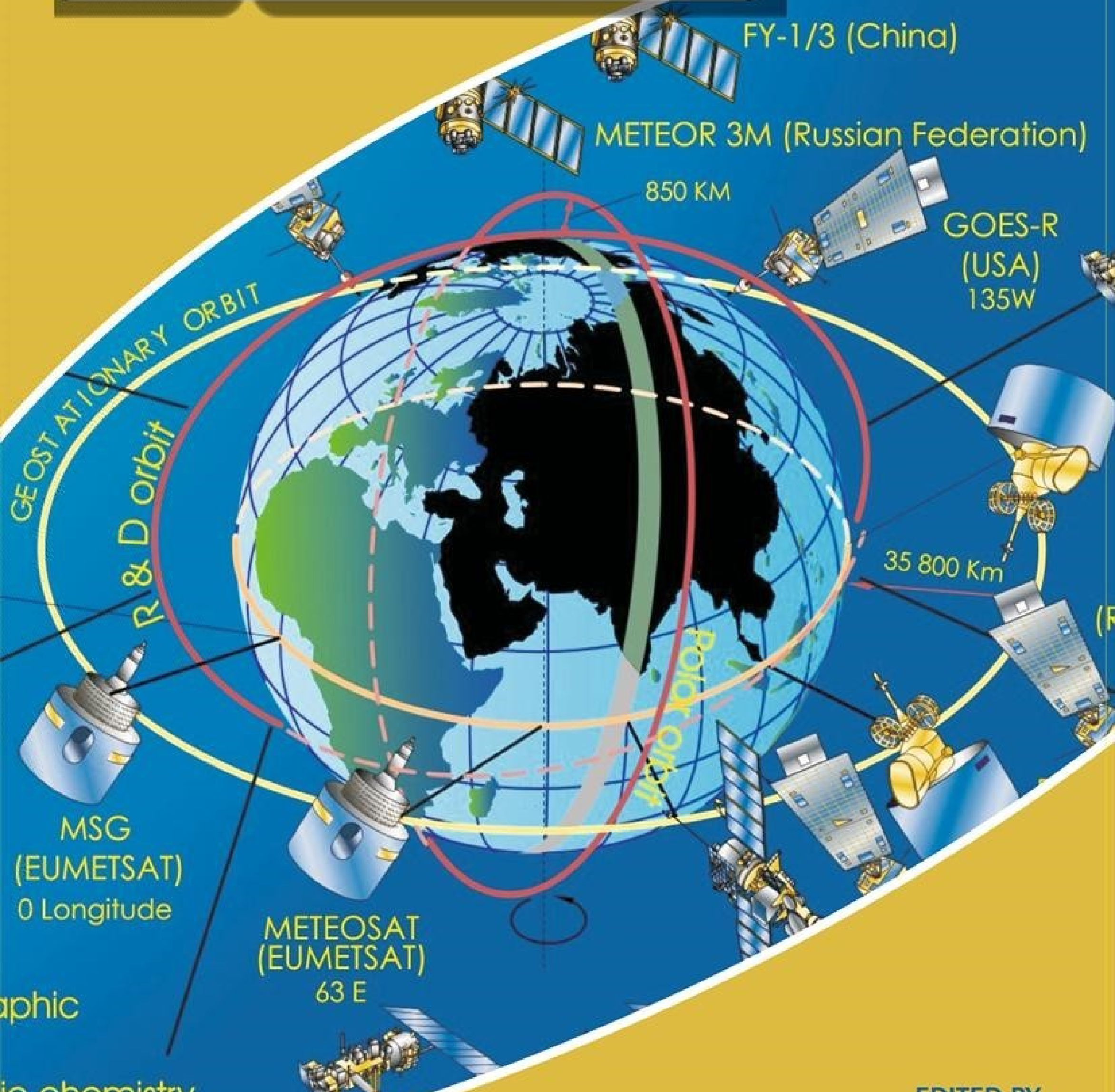
GEOSPATIAL TECHNIQUES

for Managing Environmental Resources



EDITED BY
Jay Krishna Thakur
Sudhir Kumar Singh
AL. Ramanathan
M. Bala Krishna Prasad
Wolfgang Gossel

Springer

Geospatial Techniques for Managing Environmental Resources

Geospatial Techniques for Managing Environmental Resources

Edited by

Jay Krishna Thakur

Department of Hydrogeology and Environmental Geology
Institute of Geosciences and Geography
Martin Luther University, Halle (Saale), Germany

Sudhir Kumar Singh

Centre of Atmospheric and Ocean Science
KBCAOS, IIDS, University of Allahabad, India

AL. Ramanathan

School of Environmental Sciences
Jawaharlal Nehru University, New Delhi, India

M. Bala Krishna Prasad

Earth System Science Interdisciplinary Center
University of Maryland, USA

Wolfgang Gossel

Department of Hydrogeology and Environmental Geology
Institute of Geosciences and Geography
Martin Luther University, Halle (Saale), Germany

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-94-007-1857-9 (HB)
ISBN 978-94-007-1858-6 (e-book)

Copublished by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands
with Capital Publishing Company, New Delhi, India.

Sold and distributed in North, Central and South America by Springer,
233 Spring Street, New York 10013, USA.

In all other countries, except SAARC countriesó Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lankaó sold and distributed by Springer, Haberstrasse 7, D-69126 Heidelberg, Germany.

In SAARC countriesó Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lankaó sold and distributed by Capital Publishing Company, 7/28, Mahaveer Street, Ansari Road, Daryaganj, New Delhi, 110 002, India.

www.springer.com

Cover photo credit: http://www.wmo.int/pages/prog/www/images/GOS/Figure%20II-10%20Satellites.jpg

Printed on acid-free paper

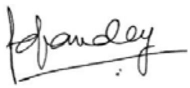
Printed in India.

Foreword

ìGeospatial informationî is spatial data concerning a place or, in space, collected in real time. Geospatial techniques together with remote sensing, geographic information science, Global Positioning System (GPS), cartography, geovisualization, and spatial statistics are being used to capture, store, manipulate and analyze to understand complex situations to solve mysteries of the universe. These techniques have been applied in various fields like metrology, forestry, air-water-land management, agriculture, health, homeland security etc. around the globe. This book has made an effort to provide a single record of current research status, technologies and applications of geospatial techniques in various fields and will be useful to academics, researchers and industry practitioners who are involved or interested in the study, use, design and development of advanced and emerging geospatial technologies around the world with ultimate aim to empower individuals and organizations in building competencies for exploiting the opportunities of the knowledge society.

The chapters in this book demonstrate the need for the integrating environmental management approaches in combination with other fields using geospatial techniques. This book presents case studies and examples from various parts of the world and provides a broad overview of various geospatial management approaches and tools used to promote environmental sustainability in the context of global environmental change and natural hazards. These case studies provide an insight into present day issues, challenges and opportunities; and highlight the key features, application of commercial and non-commercial satellite and sensors, principles and new approaches that need to be considered in future efforts for efficient integrated management in these fields. It is necessary to address the scientific strategies which underpin the implementation of geospatial techniques development for overall integrated management of various fields.

I would like to congratulate the great efforts of the editorial team of this volume which will serve as a scientific information base for future planning of the geospatial techniques development around the world and energize synergy among academicians, researchers, stakeholders, INGOs, policy makers and the corporate sector for documentation and dissemination of knowledge for the global welfare management.



Dr. Prem Chand Pandey
Emeritus Professor, Indian Institute of Technology, Kharagpur, India
Founder Director, National Centre for Antarctic & Ocean Research (NCAOR), Goa, India

Preface

The recent advancement of geospatial techniques and geodatasets have grown dramatically in size and number and become more widely distributed. The term Geospatial is generally used to describe the combination of spatial analytical methods and software with geographic datasets in conjunction with Geographic Information Systems (GIS), spatial statistics, remote sensing, cartography, geo-visualization and geomatics for aquatic and terrestrial environment especially for a civilian, business, government and military organizations. GIS as a system is integrating hardware, software, and data that captures, stores, analyzes, manages, and presents data that are linked to the locations with the merging of statistical analysis, and database technology. It is being used in cartography, remote sensing, land surveying, utility management, natural resource management, photogrammetry, geography, urban planning, emergency management, navigation, and localized search engines. Remote sensing is the acquisition of information of an object or phenomenon, by the use of either recording or real-time sensing device(s) that are wireless, or not in physical or intimate contact with the object. While, Geomatics is a fairly new terminology which is the combination of geodesy and geoinformatic terms and includes the tools and techniques used in land surveying, remote sensing, cartography, GIS, Global Navigation Satellite Systems (GNSS), photogrammetry, and related forms of earth mapping. Integration of these research and developments to real world application in these fields are now required for local, regional and global development. In this umbrella book of geospatial techniques, there are total seventeen chapters presenting the recent advancement in the GIS, remote sensing, Global Positioning System (GPS), geovisualization and geomatics for application in the field of Agriculture, Business GIS, Climate change, Energy, Environmental services, Geology, Glaciology, Health, Land information system, Natural hazard management, Natural resource management, Corporate studies, Urban planning and Utility management.

This book represents the inter- and multi-disciplinary view of authors for a meaningful and practical guidance for the better application of geospatial techniques that comprises contribution from distinguished scientists from both academia and institutions from around the globe who are actively working in this area. This volume presents the case studies and examples from various parts of the globe and provides a broader overview of various approaches involved in data acquiring, monitoring and dissemination methods. This book provides a detailed knowledge of various platforms, satellites, and sensors; tools and techniques and application to various fields for the sustainable management of environmental resources in the context of global environmental change and natural hazards.

To the beginners, we suggest not to be panicked by seeing advanced methods, applications and integrated approaches. It is necessary to understand from basics and link presented methods and application of techniques to other fields, as there are amalgamation of various processes and thoughts. The book provides a holistic application of geospatial techniques in managing natural resources in the present scenario. This book is a comprehensive collection of articles from around the world, which reflects importance of this technology in making the earth habitable. These case studies provide an insight into present-day issues, challenges and opportunities; and highlight the key features, principles, and new approaches that need to be considered in future efforts to make the earth a livable place.

We would like to thank all the contributors to this volume and take this opportunity to acknowledge all our colleagues for their time-consuming efforts to review the manuscripts of the chapters for this volume. Their efforts with high quality review of the manuscripts contributed significantly to keep the high scientific contents of the book. The editors also would like to thank collaborators and research scholars for supporting their research activities that helped in bringing out this volume successfully.

The book is scientific and research-oriented, yet appealing to a broad audience interested in environmental applications of geospatial technologies, thus making it an important reference source for the fields of GIS, remote sensing, geography, environmental policy, and environmental science. We hope that this book will be useful for geospatial scientists, remote sensing and GIS scientists, engineers, managers and administrators in both academia and industries, decision and policy makers and for government and regulatory bodies dealing with geospatial issues by providing an opportunity to acquire relevant scientific information and experiences in ì Geospatial Techniques for Managing Environmental Resourcesî .

Lastly, the editors wish to thank the publishers for bringing out this volume successfully.

Jay Krishna Thakur
Sudhir Kumar Singh
AL. Ramanathan
M. Bala Krishna Prasad
Wolfgang Gossel

About the Editors

Jay Krishna Thakur is Higrade Fellow at Department of Hydrogeology and Environmental Geology, Martin Luther University, Halle, Germany and visiting scientist at Helmholtz Centre for Environmental Research, UFZ, Leipzig, Germany. He has specialized on application of geospatial techniques in hydrogeology, hydrology, water resources and environmental management and worked extensively on the wetland management studies, quantification of water cycle components and its dynamics, modelling of surface water and groundwater, groundwater monitoring network optimization, GEONETCast environmental data streams for climate, water and environmental studies, clouds process, climate change impacts and adaptation, analysis and monitoring techniques of climate change etc. He has been awarded with various research fellowships including Netherlands Fellowship Program and Higrade Fellowship Program. He has also contributed in conference papers, book chapters and papers in reputed peer-reviewed journals.

Sudhir Kumar Singh is working at the Centre of Atmospheric and Ocean Science, University of Allahabad, Allahabad, India. He has also worked at Department of Forestry, Mizoram Central University, Aizwal, India. His main research areas are Remote Sensing and GIS applications in the field of resource management, environmental modelling and environmental pollution studies in order to solve real world problems. He has received many prestigious scholarships and has national and international collaboration with leading institutes in the United Kingdom, Germany, Turkey and Nepal. He has contributed many research papers in peer-reviewed journals and has reviewed papers, CVs and research proposal for several journals and various research institutions including the National Research Foundation, South Africa. As a supervisor, he guided over fifteen Diploma and MSc theses.

AL. Ramanathan, PhD in sedimentary geochemistry from Centre of Advanced Studies, Panjab University, worked as faculty in Geology Department, Annamalai University, for a decade. He is currently Professor of environmental geology, hydrogeochemistry, biogeochemistry, glaciology and application of geospatial techniques in environmental hydrogeology in the School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India. He has specialized on application of geospatial techniques in hydrogeology, hydrology, environmental resources management and worked extensively on the mangroves, estuaries, coastal ground waters, surface and groundwater contamination modelling of India for the past two decades. He is engaged in coastal research with a number of universities and research organisations in India, Australia, France, Japan, Sweden, Russia, USA, etc. He has guided over thirty PhDs and MPhils in the above subject and published more than seventy papers in referred reputed journals. He has also published twelve books and contributed several chapters in many others. He has thrice received the IFS Sweden (Project) Awards to work on mangrove biogeochemistry besides working in national and international projects from Indo-Australian, Indo-Russian, DST, MoEF, MoWR, etc. He has been guest editor for the International Journal of Ecology and Environmental Sciences. He is a member of editorial board in the Indian Journal of Marine Sciences and served as referee for many national and international journals as well.

M. Bala Krishna Prasad is a Research Associate at Earth System Science Interdisciplinary Center, University of Maryland, College Park, USA. He did PhD in Biogeochemistry at School of Environmental Sciences, Jawaharlal Nehru University, India. His primary research interest is evaluation and modelling of climate change and anthropogenic impacts on the water quality and aquatic ecosystem dynamics. He is a recipient of the Young Scientist Award from the Ministry of Science & Technology (India), Linnaeus-Palme Research Fellowship from Department of Earth Sciences, Uppsala University (Sweden), Berkner Fellowship of AGU (USA) and Junior Research Fellowships as well. He has contributed more than a dozen papers in reputed peer-reviewed international journals.

Wolfgang Gossel, habilitated in 2008 in Applied Geosciences at Martin Luther University, Halle (Germany) and worked for nearly 20 years in hydrogeology. He is a specialist in groundwater flow and transport modelling, GIS data processing and the application of interpolation techniques in hydrogeology. He is mainly working on the regional scale and since 2002 has got published several papers about the numerical groundwater flow models of the Nubian Aquifer System in Northeast Africa and the map Hydrogeology of Germany. He has also collaborated with the Jawaharlal Nehru University in New Delhi, Assiut University and Ain Shams University in Egypt, Universidad Autonoma de Merida and Universidad Autonoma de San Luis Potosi in Mexico. He was Associate Editor of Hydrogeology Journal and Grundwasser. As a supervisor he guided about 50 Diploma and MSc theses.

Contents

1

Environmental Informatics: Advancing Data Intensive Sciences to Solve Environmental Problems

Chaowei Yang, Yan Xu[1] and Daniel Fay[1]
Center for Intelligent Spatial Computing, and Department of Geography and GeoInformation Sciences, George Mason University
Fairfax, VA, USA
[1]Earth, Energy and Environment at Microsoft Research Connections
Microsoft Corporation, Redmond, WA, USA

Introduction

The 21st Century witnesses emergence of geospatial cyberinfrastructure and other relevant geospatial technologies (Yang et al., 2010) for collecting data, extracting information, simulating phenomena scenarios, and supporting decision making (Caragea et al., 2005; Stadler et al., 2006). The advancements of the geospatial technologies not only provide great opportunities for us to better understand environmental issues and better position us to solve global to local environmental problems (Pecar-Ilic and Ruzic, 2006), but also pose great challenges for us to handle terabytes to petabytes of heterogeneous environmental data. Environmental informatics (Green and Klomp, 1998; Hilty, Page and Hŗebíçek, 2006) should be revisited to efficiently and effectively manage, integrate, and mine information and knowledge from the vast amount of data for supporting environmental decisions (Hey, Tansley and Tolle, 2008).

Information/knowledge mining often starts with heterogeneous data from various resources, and ends with re-aggregated data in numbers and/or intuitive graphics adopted by decision makers. On one hand, there are many conventional geospatial techniques that can be used to facilitate generating information and knowledge from environmental data to produce social impact. There are also many computing technologies that are not specifically designed for managing environmental data but can be used innovatively to significantly accelerate the mining process. On the other hand, it is a challenge to decide what techniques to choose and to integrate with, so that strategies could be taken to

J.K. Thakur et al. (eds.), *Geospatial Techniques for Managing Environmental Resources*,
DOI 10.1007/978-94-007-1858-6_1, © Capital Publishing Company 2012

design solution by leveraging the techniques efficiently in managing environmental data.

Based on our past decade's investigation on utilizing geospatial technologies for managing environmental data, this chapter illustrates how to utilize a set of computational geospatial techniques to manage and process environmental data through presenting five examples. The chapter concludes with a discussion on future directions of environmental informatics to handle potential problems from the arena of data intensive, computing intensive, spatiotemporal intensive and concurrent access intensive (Yang et al., 2011).

Background

Environmental data are of essential values in support of environmental decisions by providing historical and real-time representation of the facts about environmental phenomena. The data have the characteristics of: (1) global distribution (Pillmann, Geiger and Voigt, 2006; Zhizhin et al., 2007), for example, dust storm generated from northern Africa could travel to Europe and forest fire emissions in Washington state can impact air quality in Iowa (Xie et al., 2010); (2) multi-dimensionality, though many datasets are two dimensional, the collection of three and four dimensional data has begun, for example, a four dimensional dataset is measured of daily nitrogen concentrations at different depths within Chesapeake Bay; (3) heterogeneity, which exists in many ways such as within data formats, data contents, data accuracy, data spatial resolution and time scale (Devarakonda et al., 2010; Yang et al., 2008), geographic coverage, and jurisdiction or culture boundaries; and (4) large volume, intensive data up to petabytes in size may be collected to record observations in four dimensions of space and time (Goodchild, Yuan and Cova, 2007).

The emerging requirements for data of high resolution and collected as a time series when compared to existing information may result in gaps in time, geographic coverage, and jurisdiction or culture boundaries. The time scales over which relevant environmental data may be collected also varies by the type of observations that are required (Karatzas, Nikolaou and Moussiopoulos, 2004), for example, data in response to environmental disasters, climate change and/or policy decisions may require daily, yearly or decadal measurements. Spatial constraints and scientific principles must also be represented within environmental datasets (Yang et al., 2011b). An illustration of this is how toxic material will travel in rivers at a much faster speed than hazardous material percolating through soil, but slower than hazardous gas emissions travelling through the atmosphere. These movements of material are governed by different geophysical principles and therefore require different models of various complexities to simulate and produce accurate predictions in support of decision making.

From an informatics perspective, these problems are also reflected as data intensive, computing intensive, spatiotemporal intensive, and concurrent access intensive (Yang et al., 2011a): (1) the data in terabytes to petabytes are volume intensive; (2) the various geophysical models and interpolation methods require cutting-edge computer or high end computing support to simulate environmental phenomena; (3) data are increasingly demanded to be tagged up with both spatial and temporal information in order to be useful (Rasuly, Naghdifar and Rasoli, 2010). Near-real time data requirements are becoming more common for environmental emergency responses that happened much more frequently within our globalized world (Karatzas et al., 2004); and (4) some datasets and model predictions are of public interest to be accessed by thousands to millions of citizens in a very short time period. For example, the air quality information, just like the weather information, is essential for the public to make personal decisions about their outdoor activities and travel plans based on individual health considerations.

Best handling environmental data with such characteristics and satisfying the requirements from the users require us to revisit environmental informatics, the science studying how to utilize information technologies and methodologies to process environmental data for better informed decision making (Wang 2007). In general, environmental informatics includes the procedure (Fig. 1) to: (1) extract, interpolate, and aggregate data to generate information for specific environmental decision needs; (2) analyze or simulate environmental issues to obtain knowledge about environmental issues; (3) support specific environmental decisions with potential scenarios and domain knowledge; and (4) making social impact by contributing valuable information and collecting citizen feedbacks to improve the environmental observations, data processing techniques, model simulation algorithms, or to revise decision making procedure with a more scientifically sound approach.

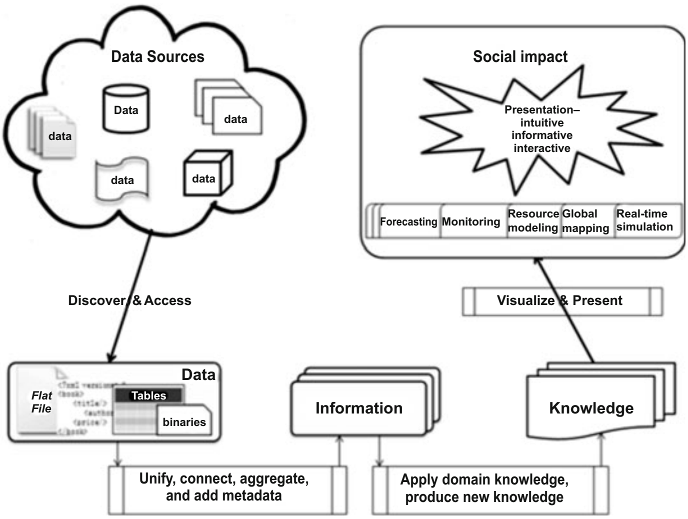


Figure 1. Information flow within environmental informatics.

Research Examples

Observing the characteristics of environmental data, the requirements for data processing, and the procedure to address user requirements, we review environmental informatics in the context of data intensive science that includes transforming environmental data, to information, to knowledge, to decision support, and ultimately to social impact. Each of the four steps is exemplified through literature reviews and research examples from the Environmental Informatics Framework, a Microsoft eScience initiative and George Mason University (GMU) Center for Intelligent Spatial Computing (CISC)'s research projects: (1) A virtual Earth observatory project shows the discovery and integration of data from widely distributed services for environmental operations based on interoperability and on-the-fly integration; (2) A United States ozone observation and interpolation project illustrates data to information transformation; (3) An on-demand MODIS science project demonstrates how cloud computing can help to enable science question answered in near real-time; (4) A Chinese land cover change project demonstrates how to utilize intensive data and sophisticated models to support decision making; and (5) A citizen sensing flooding project is used to demonstrate how the data, information, modelling, and knowledge can produce social impact and benefit from citizen sciences. Scenarios with integrative solutions based on cutting-edge Microsoft data and information technologies are laid out in the examples to illustrate the future of environmental informatics.

Data Discovery, Access and Utilization with Virtual Observatory

Global Earth observations collect petabytes of geospatial data that are important for tackling environmental issues. But the datasets are widely distributed, in diverse volume, and served through heterogeneous approaches. Efficient discovery, access and utilization of the data are great challenges. The efficient integration of the data on the fly will provide enabling capability to facilitate global to local level decision support (Yang et al., 2011b). Under the sponsorship from Microsoft and NASA, GMU CISC developed a spatial web portal to integrate the data services from around the global to serve as a virtual Earth observatory.

Figure 2 illustrates the interface of the spatial web portal, which is able to search against global Earth observation data in collaboration with the clearinghouse of Global Earth Observation System of Systems (GEOSS). Found datasets can be accessed and integrated from around the world (denoted red icons) when served through interoperable interfaces, such as Open Geospatial Consortium (OGC)'s Web Services, including Web Map Service (WMS), Web Coverage Services (WCS), and others.

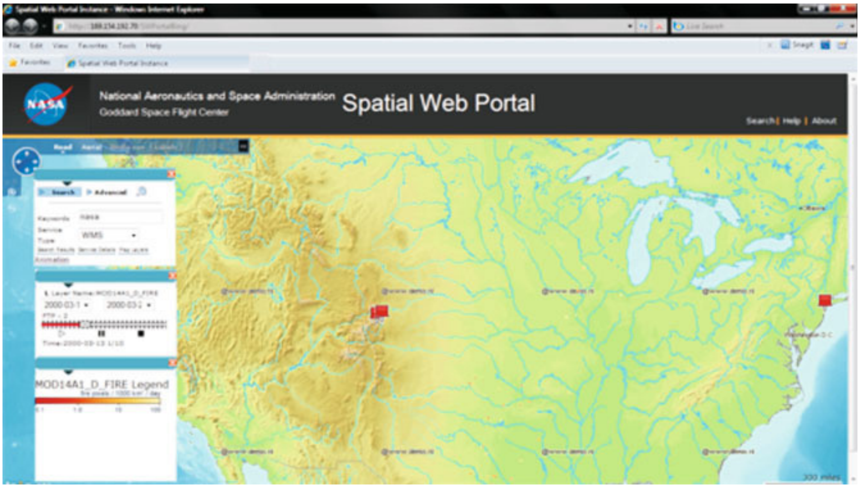


Figure 2. A virtual observatory of Earth science data based on spatial web portal.

Figure 2 illustrates that forest fire information, detected through MODIS, is integrated on a daily basis through the spatial web portal and decision makers can, for example, use it to generate historical animations (controlled by the three small windows) as well as ingest real-time Earth observations to assist emergency response for allocating fire suppression resources. The system can also be used for other environmental problems because of its characteristics of interoperability, and the advanced capabilities provided by Microsoft Bing Maps.

Ozone Observation and Country-wide Interpolation

Ozone concentration is an important air quality factor that has impact on our breathing and other public health related activities. In collaboration with NOAA, EPA established about 100 observation sites around the United States. These sites generate point data on a minute basis and most of them have time series observations. With the outage variations and the uneven distribution of the sites (denoted in dots in Fig. 3), it is difficult to get an overview about the ozone distribution around the entire country with the point data.

Starting from a course project, a team of students at GMU collected daily data from the sites for the past 10 years and developed a graphics user interface (Fig. 3) for end users to: (1) pinpoint the ozone concentration within the map, (2) choose the specific dates for the observation (through upper left calendar), (3) animate the ozone concentration on a daily basis, and (4) choose different interpolation methods and accuracy to interpolate the point data to get area coverage for the entire nation (lower left).

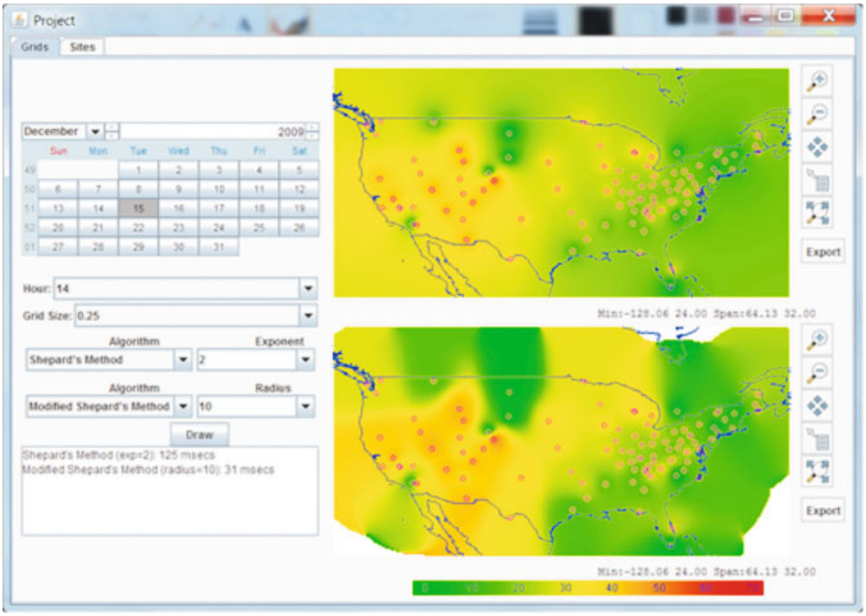


Figure 3. Ozone interpolation and navigation tool for the United States.

This tool could help us to have a better view of the entire nation for ozone concentration based on the site point observations and statistical interpolation. To some degree, it helps us to interpolate or extract information for missing places. But on the other hand, it demonstrates, through comparison of interpolation methods (right side images of Fig. 3), that simple interpolation is too arbitrary to generate accurate ozone concentration maps. Therefore, follow-on studies are needed to incorporate weather information and other source of observations, such as satellite remote sensing, to generate more accurate interpolations and eventually to forecast the ozone concentrations based on the historical and weather forecasting (Slini, Karatzas and Moussiopoulos, 2003). Making the data and interpolation services available through OGC web services would also help them to be easily discovered, accessed and utilized through, for example, virtual observatory described in previous section.

Knowledge Generation from Long-term Observation

Earth observations are invaluable for scientists to track the phenomena facts of the Earth surface. Complex data processing algorithms, such as subsetting and parameter extracting, are needed to process the observations before final products can be used by scientists or decision makers. And these products requirements are quite different. Therefore, an infrastructure is needed to provide on-demand science queries for specific parameters or products.

As an example of real-time data intensive project where not only large amount of data but also complex processing algorithms are needed, the Microsoft Research MODISAzure project demonstrated that cloud computing technology can dramatically change the scale and the pace of the environmental research. It enables scientists to ask more exciting and important questions and will help to answer them by providing an automated closed processing loop to support the observation, data selection, reprojection, derivation reduction, and analysis reduction for science results given specific scientific questions (Fig. 4).

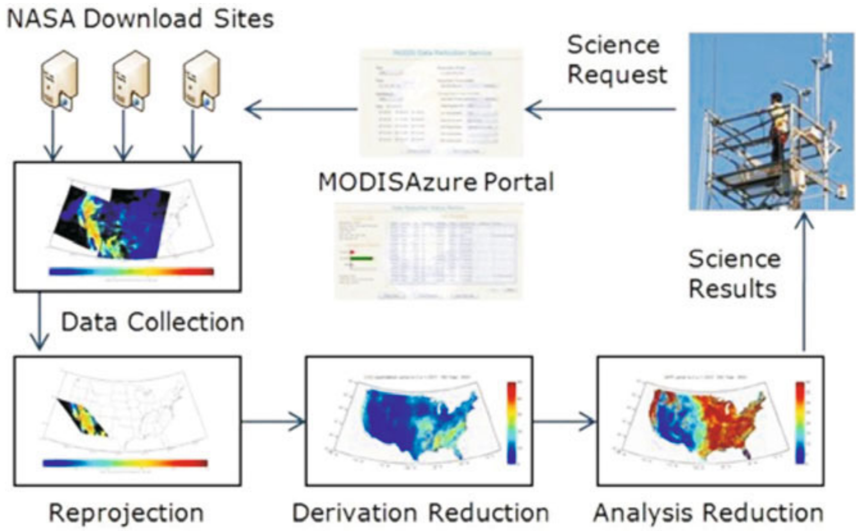


Figure 4. The workflow of MODISAzure is implemented on the Azure cloud computing platform.

This science on-demand scenario includes data subsetting/selection, reprojection, derivation and analyses components based on complex algorithms, which are data intensive and computing intensive. Spatiotemporal intensive is also a characteristic that is considered within the project to provide real time response for science questions. The data intensive, computing intensive and spatiotemporal intensive issues are resolved within this project using Azure cloud computing platform.

The solution automates the main components of the pipeline including data download, image reprojection, on-demand execution of the algorithms, and delivery of final results to scientists' desktops. It helps solve several obstacles to accessing the vast and varied remote sensing data from the MODIS satellites and other sources. More details about the MODISAzure project can be found at *http://research.microsoft.com/projects/azure/azuremodis.aspx*. This project demonstrates that the data discovery, access and utilization can be conducted in an automated fashion with support from cloud computing.

Forecasting Changes in Urban Land Use

Land use/cover change is a driving factor for many environmental issues. Forecasting the environmental consequence due to fast urban expansion is a serious challenge for decision makers in many fast developing countries. A capability to simulate/forecast land use and land cover change based on specific policies and constraints is important to make informative decisions (Rasuly, Naghdifar and Rasoli, 2010).

Figure 5 shows a screen shot of an online Bing Maps based land use application. High Performance Computing (HPC) technologies are applied to combine urban land expansion module and environmental forecast module to provide computing support. It enables real-time simulation of the effect selected by heterogeneous environmental variables (shown on the upper-right corner in the picture). The application provides an interactive user interface for decision makers and the general public to investigate and better understand how the changes in urban land use may impact the environment. The platform is accessible via Internet browser with user access control. The application can be easily scaled to incorporate more environmental factors and regions.

The project integrates historical and observation data from different resources including in-situ sensor and satellite images. In addition, land use and land change knowledge based on different driving factors, such as policy constraints and spatial distribution, are built into a model that is running on a high performance computing facility at Tsinghua University.

More detailed description on this application can be found in the proceedings of Microsoft Environmental Research Workshop 2010 (*http://research.microsoft.com/en-US/events/environmentalresearch2010/default.aspx*)

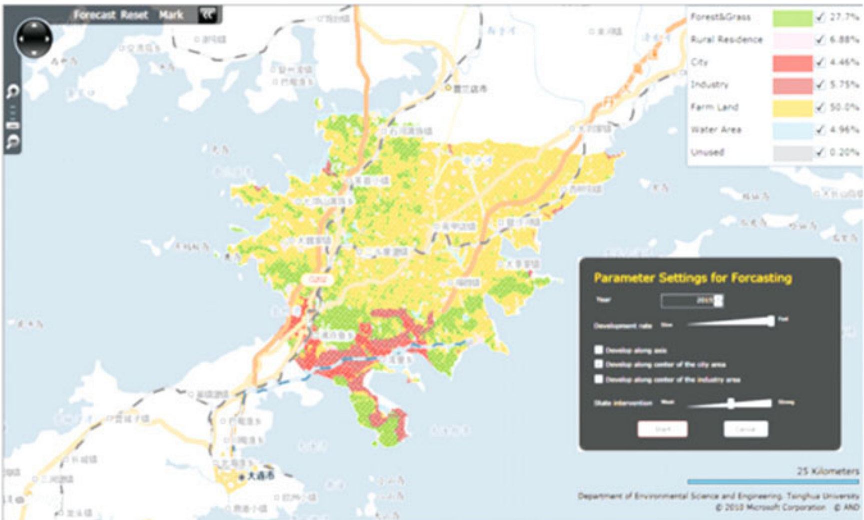


Figure 5. The land use change simulation based on different parameters setting.

Social Impact and Citizen Sensing for Flooding Management

Many environmental issues have personal touch and broad impact. It is challenging to integrate information from masses of individuals within a broad area impacted. For example, in regions such as south Florida, before a hurricane hits, the city receives large number of calls from residents to report house flooding. Such citizen call-in data are location and time sensitive during the emergency, which can be leveraged to democratize emergency management, flooding control and decision support; raise localized, timely and socially relevant situational awareness; and critically complement physics-based modelling forecasts (Barros, 2005). A better visualization of the call-in data in time series will help decision makers or emergency planners to manage emergency preparedness. In this case, Microsoft Research World Wide Telescope (WWT, http://www.worldwidetelescope.org) is found to be the ideal tool to produce the desired visualization effect. Figure 6 shows a screen shot of the citizen call-in data in 2005-2008 animated in a WWT guided tour with different colours denoting different years. The tour can be interactively enriched with scientific models to simulate and forecast potential storm hits.

Figure 6. Citizen sensing helped with the flooding management of southern Florida.

Solutions with Microsoft Advanced Technologies

Through five examples, previous section illustrates how Microsoft technologies are utilized to solve data intensive, computing intensive, spatiotemporal intensive and concurrent access intensive issues facing us for advancing environmental informatics in the 21st century.

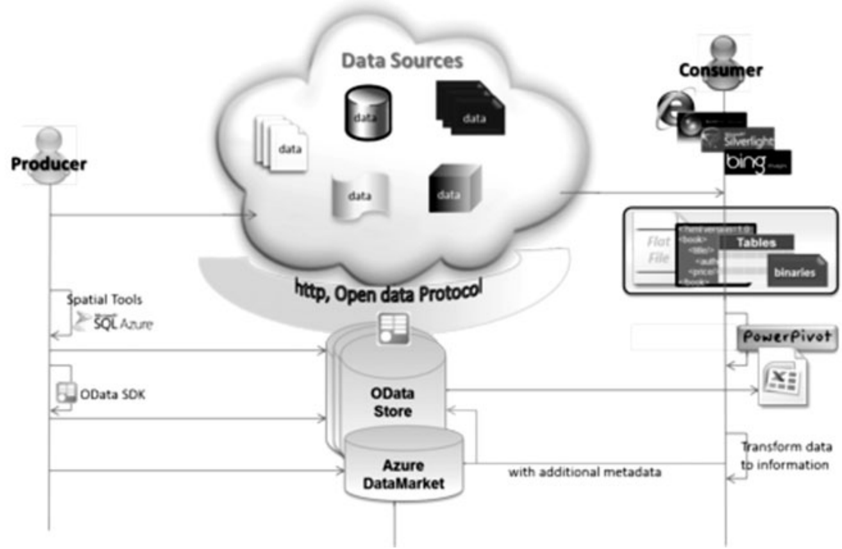


Figure 7. A flexible solution from Environmental Informatics Framework (EIF), a Microsoft eScience initiative.

Figure 7 illustrates a flexible configuration and possible components that can help environmental data processing and advancing environmental informatics. The solution is focussed on data generated from a variety of resources and produced by different users. Microsoft SQL and SQLAzure platform can be used to organize and manage the data kept in OData Store and DataMarket. OData Stores and Microsoft Azure DataMarket (formally codenamed Dallas) could provide connection to the data through the Open Data Protocol (OData), and consequently visualized in a variety of tools, such as Microsoft Excel, Excel with PowerPivot Add-in, Bing Maps, or WWT, for end users to intuitively access the datasets. If all components were running in a cloud computing environment, such as Azure, the data, services utilized can be discovered and utilized in an automatic fashion with minimum scripting development.

Future Research Directions

Advanced IT technologies are providing enabling capabilities to better tackle environmental issues in the 21st century as illustrated through the examples in an environmental informatics fashion (Radermacher, 1994). Problems also emerged from the process of the data accumulation, scientific research, end user requirements, and requirements for more precise information and knowledge. The challenges of data intensive, computing intensive, spatiotemporal intensive, and concurrent intensive will continue to serve as driving demands for advancing environmental informatics. Opportunities and challenges coexist within our efforts to tackle the issues (Yang et al., 2011b). The following directions need further attention for research and development in the following decade to advance environmental informatics.

- An environmental informatics *infrastructure* is needed for fast integration and aggregation of environmental observations from different organizations into a virtual simulation centre that can produce scientifically sound decision support information (MacDonell, Morgan and Newland, 2002). This can benefit from the recent geospatial cyberinfrastructure advancement at different agencies (Yang et al., 2010), especially within the environmental protection agency and companies, such as Microsoft, who are willing to contribute advanced technologies to address our global environmental issues.
- Advanced *computing technologies* are still needed to be further developed, especially in a spatial cloud computing fashion (Yang et al., 2011a and 2011b) to best leverage the distributed computing resources to handle environmental observations accumulated from local to global locations and to best handle environmental models, observations, and simulations in an optimized fashion.
- *Data integration and interoperability* is an inevitable task that will endure the entire life cycle of environmental informatics to solve data heterogeneous and service heterogeneous problems in the 21st century.
- *Spatiotemporal models* are becoming increasingly important from both theoretical and technological research and development to handle real-time and historical environmental data for better simulating potentially resulting scenarios from environmental decisions.
- *Human knowledge* about our planet is critical to the advancement of environmental informatics (Chen et al., 2007) and should be integrated in solutions (Tochtermann and Maurer, 2000). But our understanding of Earth has been very much divided according to different domains including, for example, biology, physics, ocean, atmosphere and land. *Model integration* is much needed to couple individualized models to form integrative systems to better simulate the response of geophysical systems from any event triggered by either human activities, such as hazardous material emission, or by nature, such as a tsunami.
- All resources supporting the advancements of environmental informatics will have their own relevant *quality and usage* limitation based on the providers, operational status and end users. The research on the quality and authority of data, service, information, knowledge and models will be critical to guarantee better decision making. Assessing the accuracy is essential within environmental informatics processes (Kalapanidas and Avouris, 2003).
- *Citizen science* is emerging as a phenomenon to involve citizens in popular scientific discoveries and science and policy comment and verification (Gruiz, 2009). For environmental science, this is especially true when

observations are still the best approach to obtain environmental information (Mayfield, Joliat and Cowan, 2001). Large amount of environmental issues are reported by citizen (Stockwell et al., 2006). Citizens could also contribute their computing resources in a Seti@Home fashion to solve public environmental issues, such as climate change through Climate@Home.

- *An automated and self-adaptive environment* for discovery, access and utilization of data and services will be needed in the future for quickly integrating data, information, knowledge, citizen sciences, and decisions made to produce knowledge for rapid, broad and precise environmental problems solving.

Conclusion

Within the data explosion era, environmental informatics faces great challenges to better manage, integrate, simulate and forecast environmental information to extract knowledge for better environmental decisions. This chapter summarizes environmental informatics through analyzing the characteristics of environmental data and five research examples.

We report: (1) How the environmental data observed can be interpolated to have a reasonable geographic coverage; (2) How data can be discovered, accessed and integrated and visualized through Bing Maps to conduct information integration on the fly; (3) How to utilize Azure cloud computing and MODIS knowledge to produce decision support results in real time; (4) How to utilize cloud computing and distributed data integration to support environmental scenario simulations for decision making with potential scenarios; and (5) How to involve citizens in environmental sciences to help observe, respond and mitigate hazardous issues. Through advanced technologies provided by Microsoft, these projects demonstrate exemplary solutions for the four steps within environmental informatics.

The research benefits from recent advancements in computer science, data sciences, and information technologies and methodologies. To fully integrate observation data worldwide, utilize elastic cloud computing platforms, and serve users with right environmental information in the right place, at the right time will be a driving objective for the next decade in advancing environmental informatics for (1) solving environmental issues, (2) advancing environmental sciences, (3) responding to environmental emergency, and (4) protecting our home planet for future generations.

Research is needed in (1) bridging advanced technologies and environmental sciences, (2) advancing sciences and technology, information theory within a globalized context, and (3) improving end users awareness and utilization of environmental data, information and knowledge in a more creative and beneficial decision environment.

Acknowledgements

Research reported is sponsored by Microsoft, NASA and Federal Geographic Data Committee. The authors would also like to acknowledge Yi Liu from Tsinghua University, China, for his contribution in the example as shown in Fig. 5; South Florida Water Management District, USA, and Yong Liu from National Center for Supercomputing Applications (http://www.ncsa.illinois.edu/), for their contribution in the example as shown in Fig. 6; Xi Zhou, Chen Xu, Jizhe Xia, Xin Qu, and Min Sun, Qunying Huang from George Mason University (cisc.gmu.edu) for their contributions to examples as shown in Figs 2 and 3.

References

Barros, A.P. (2005, Jul. 31-Aug.4). Environmental informatics - Long-lead flood forecasting using Bayesian neural networks. Paper presented at the International Joint Conference on Neural Networks, Montreal, Canada.

Caragea, D., Zhang, J., Bao, J., Pathak, J. and Honavar, V. (2005). Algorithms and software for collaborative discovery from autonomous, semantically heterogeneous, distributed information sources. *Lecture Notes in Computer Science*, **3734**, 13–44.

Chen, Z., Gangopadhyay, A., Karabatis, G., McGuire, M. and Welty, C. (2007). Semantic integration and knowledge discovery for environmental research. *Journal of Database Management*, **18**, 43–68.

Devarakonda, R., Palanisamy, G., Green, J.M. and Wilson, B.E. (2010). Data sharing and retrieval using OAI-PMH. *Earth Science Informatics*, **3**, 1–5.

Dhanushkodi, S.R., Mahinpey, N., Srinivasan, A. and Wilson, M. (2008). Life cycle analysis of fuel cell technology. *Journal of Environmental Informatics*, **11**, 36–44.

Goodchild M., Yuan, M. and Cova, T.J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science*, **21**, 239–260.

Green, D.G. and Klomp, N.I. (1998). Environmental informatics—A new paradigm for coping with complexity in nature. *Complexity International*, **6**.

Gruiz, K. (2009). Web-based information system and decision support tool: The structure and use of the MOKKA IT tool. *Land Contamination and Reclamation*, **17**, 695–702.

Hey, T., Tansley, S. and Tolle, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Press, Redmond, WA.

Hilty, L.M., Page, B. and Hṛebíçek, J. (2006). Environmental informatics. *Environmental Modelling and Software*, **21**, 1517–1518.

Karatzas, K., Nikolaou, K. and Moussiopoulos, N. (2004). Timely and valid air quality information: The APNEE-TU Project. *Fresenius Environmental Bulletin*, **13**, 874–878.

Kalapanidas, E. and Avouris, N. (2003). Feature selection for air quality forecasting: A genetic algorithm approach. *AI Communications*, **16**, 235–251.

MacDonell, M., Morgan, K. and Newland, L. (2002). Integrating information for better environmental decisions. *Environmental Science and Pollution Research*, **9**, 359–368.

Mayfield, C., Joliat, M. and Cowan, D. (2001). The roles of community networks in environmental monitoring and environmental informatics. *Advances in Environmental Research*, **5**, 385–393.

Pecar-Ilic, J. and Ruzic, I. (2006). Application of GIS and Web technologies for Danube waterway data management in Croatia. *Environmental Modelling and Software*, **21**, 1562–1571.

Pillmann, W., Geiger, W. and Voigt, K. (2006). Survey of environmental informatics in Europe. *Environmental Modelling and Software*, **21**, 1519–1527.

Radermacher, F.J., Riekert, W.-F., Page, B. and Hilty, L.M. (1994). Trends in environmental information processing. *IFIP Transactions. A: Computer Science and Technology* **(A-52)**, 597–604.

Rasuly, A., Naghdifar, R. and Rasoli, M. (2010). Detecting of Arasbaran forest changes applying image processing procedures and GIS Techniques. *Procedia Environmental Sciences*, **2**, 454–464.

Stadler, M., Ahlers, D., Bekker, R.M., Finke, J., Kunzmann, D. and Sonnenschein, M. (2006). Web-based tools for data analysis and quality assurance on a life-history trait database of plants of Northwest Europe. *Environmental Modelling and Software*, **21**, 1536–1543.

Slini, T., Karatzas, K. and Moussiopoulos, N. (2003). Correlation of air pollution and meteorological data using neural networks. *International Journal of Environment and Pollution*, **20**, 218–229.

Stockwell, D.R.B., Beach, J.H., Stewart, A., Vorontsov, G., Vieglais, D. and Pereira, R.S. (2006). The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological Modelling*, **195**, 139–145.

Tochtermann, K. and Maurer, H. (2000). Knowledge Management and Environmental Informatics. *Journal of Universal Computer Science*, **6**, 517–536.

Wang, X. (2007). Environmental informatics for environmental planning and management. *Journal of Environmental Informatics*, **9**, 1–3.

Xie, J., Yang, C., Zhou, B. and Huang, Q. (2010). High performance computing for the simulation of dust storms. *Computers, Environment, and Urban Systems*, **34**, 278–290.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Fay, D. and Bambacus, M. (2011a). Spatial Cloud Computing – How geospatial science use and help to shape cloud computing. *International Journal of Digital Earth*, **4**, 305–329.

Yang, C., Raskin, R., Goodchild, M.F. and Gahegan, M. (2010). Geospatial Cyberinfrastructure: Past, Present and Future. *Computers, Environment, and Urban Systems*, **34**, 264–277.

Yang, C., Wu, H., Huang, Q., Li, Z. and Li, J. (2011b). Utilizing spatial principles to optimize distributed computing for enabling physical science discoveries. *Proceedings of National Academy of Sciences*, doi: /10.1073/pnas.0909315108.

Zhizhin, M., Kihn, E., Redmon, R., Poyda, A., Mishin, D., Medvedev, D. et al. (2007). Integrating and mining distributed environmental archives on Grids. *Concurrency Computation Practice and Experience*, **19**, 2157–2170.

2

Hierarchical Geospatial Computing Environment for Data-intensive Geographic Process Simulation

Mingyuan Hu, Hui Lin, Bingli Xu[1], Ya Hu[2], Sammy Tang[3] and Weitao Che

Institute of Space and Earth in Information Science, the Chinese University of Hong Kong, China

[1]Department of Information Engineering, The Academy of Armored Forces Engineering, Beijing 100072, China

[2]Remote Sensing Information Engineering Department, Faculty of Geosciences and Environmental Engineering Southwest Jiaotong University, Chengdu 610031, China

[3]Information Technology Services Center, the Chinese University of Hong Kong, China

Introduction

Geographic information system (GIS) professionals recognize that geographic process is essential for understanding what is happening in the world, learning how the environment is changing, modelling how complex systems work and giving context to other types of data. Consequently, geographic process models have been increasingly featured for the next generation geographic information science (system), as a method for phenomena simulation and mechanism analysis of the physical environment and its live activities, thereby driving conventional GIS based on data manipulation into the world of dynamic and computational processes (Goodchild, 2006; Yuan and Hornsby, 2008; Torrens, 2009; CSDGS et al., 2010).

However, geographic process models such as atmospheric models generally are toward computational models, large data-intensive sets based, and at multiple spatiotemporal scales, which often take time-consuming computing or need immediately simulated responses to explain these dynamic phenomena (Openshaw, 2000; Guan et al., 2006; Brimicombe, 2009). In order to make better and full use of the vast geographic data and exploit the

J.K. Thakur et al. (eds.), *Geospatial Techniques for Managing Environmental Resources*,
DOI 10.1007/978-94-007-1858-6_2,

valuable information beyond the original data resources, data intensive computing is now considered as the "fourth paradigm" in scientific discovery after theoretical, experimental and computational science (Tony et al., 2009). New computing technologies such as cloud computing, grid technology and multicore processors have spawned GIS fields to meet the new scientific challenges, including computationally-intensive scientific, mathematical and academic problems through volunteer computing (Foster, 2002; Goodchild, 2010).

Some pilot studies in geospatial fields are spatial statistical analysis (Wang et al., 2008; Harris, 2010), air-borne laser scanning data sequential processing (Han et al., 2009), hyperspectral image processing (Plaza et al., 2006), parallel cellular automata modelling (Guan and Clarke, 2010; Li et al., 2010), as well as simulations with parallelized mesoscale environmental models (Xie et al., 2009; Xu et al., 2010) and wildland fire behaviour simulation (Douglas and Coen, 2010). Despite the existence of commercially available dispersed computing resources, they are highly priced and are largely inaccessible to general users (Sugato, 2005; Zhang, 2010). Furthermore, these systems cannot provide the entire solution in their generic forms (Foster and Kesselman, 2004) and they are also difficult for geographical tasks schedule, leveraging the built-in geospatial data management as well as inherent implementing computing tasks in heterogeneous environments (Curran and Shearer, 2009; Huang et al., 2009). On the other hand, when the number of processors go beyond a threshold, normalized speedups of grid/cloud computing could slow down significantly, which might be caused by the combined factors of communication/scheduling overheads, workload imbalances and contention of shared resources (Xie et al., 2009; Li et al., 2010). There is also an increasing recognition that the scalability and controllability of high PC clusters allow progress beyond remote computational resources limited by the network bandwidth, security authentication and queue requests, etc. (Marín et al., 2011). The low cost high performance PCs thus has fostered research to use them to drive immediately geographic process computing (Zhang, 2010). In particular, personal PC clusters computing advances the linkage across existing resources, but also offers the power of self-control to the management of geospatial data resources and dataflow for real time dynamic simulation at the local area. Consequently, in this chapter, a distributed problem-solving environment is proposed in order to reveal fully the real features of spatiotemporal geographic process and support both 3D analysis and hierarchical emergency decision making, comprising the geographic process models, the hierarchical geospatial computation involving grid computing layer and local PC clusters effectively coordinated, and Viz Wall for real-time data visualization and geographic analysis.

The remainder of this chapter is organized as follows. Second part describes the related background of geographic process models and high performance hierarchical geospatial computing environment for data-intensive computing. A hierarchical geospatial computing environment for data-intensive geographic process simulation is introduced in third part. The fourth part discusses the case study: the grid computing based regional atmospheric simulation, PC cluster based parallel computation of a Gauss dispersion model, and the visual wall for geographic visualization. Final part gives the conclusion of this work with thoughts on the future prospects.

Advanced Features in Geographic Process Computation

Geographic Process Models

Geographic process models are not only geo-knowledge in them, but can also generate new geo-knowledge by changing the input initiate parameters. Therefore, being different from digital representation such as geometric construction, what a geographic process model highlights are dynamical geographical process simulation and geographical phenomenon analysis, so it focusses on not only the cities or communities, but also all geographical environments such as atmospheric pollution, oceans dynamics and soil horizon. In this case of air pollution simulation research, Mesoscale Model5 (MM5) and Gauss plume model for simulation and analysis are typically introduced.

MM5 (Dudhia et al., 2005)

The PSU/NCAR mesoscale model5 (known as MM5) is a regional, nonhydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale atmospheric circulation. The model is supported by several pre- and post-processing programs, which are referred to collectively as the MM5 modelling system. The MM5 modelling system software is mostly written in Fortran, and has been developed at Penn State and NCAR as a community mesoscale model with contributions from users worldwide. Since that time it has undergone many changes designed to broaden its applications. These include (i) a multiple-nest capability, (ii) nonhydrostatic dynamics, (iii) a four-dimensional data assimilation (Newtonian nudging) capability, (iv) increased number of physics options, and (v) portability to a wider range of computer platforms, including OpenMP and MPI systems. There are about eight modules for MM5, which are TERRAIN, REGRID, LITTLE_R/RAWINS, INTERPF, INTERPB, NESTDOWN, MM5 and GRAPH/RIP.

Gauss Plume Model (Turner, 1994)

Based on statistic theory, in steady turbulence environment, particle dispersion and transport behaviours accord with Gaussian plume model, which is also called Gaussian dispersion model. Gaussian plume model does not include chemical reaction among pollutants and wet and dry deposition, and is an idealized simplified dispersion model. Gaussian plume model is relatively