

A GEOINFORMATICS APPROACH TO LIDAR DATA DISTRIBUTION AND
PROCESSING WITH APPLICATIONS TO GEOMORPHOLOGY

by

Christopher J. Crosby

A Thesis Presented in Partial Fulfillment
Of the Requirements for the Degree
Master of Science

ARIZONA STATE UNIVERSITY

August 2006

A GEOINFORMATICS APPROACH TO LIDAR DATA DISTRIBUTION AND
PROCESSING WITH APPLICATIONS TO GEOMORPHOLOGY

by

Christopher J. Crosby

has been approved

July 2006

APPROVED:

_____, Chair

Supervisory Committee

ACCEPTED:

Department Chair

Dean, Division of Graduate Studies

ABSTRACT

The emergence of new digital data acquisition technologies in the geosciences has exciting implications for the types of data that are now available to researchers.

However, along with these datasets comes an increase in the volume and complexity of scientific data that must be efficiently managed, archived, distributed, processed and integrated in order for it to be of use to the scientific community. The rapid growth of LiDAR (Light Distance And Ranging, a.k.a. ALSM (Airborne Laser Swath Mapping)) for geoscience applications is an excellent example of the opportunities and challenges presented by these types of datasets. Capable of generating digital elevation models (DEMs) more than an order of magnitude more accurate than those currently available, LiDAR data offer geomorphologists the opportunity to study the processes that shape the earth's surface at resolutions not previously possible yet essential for their appropriate representation.

Unfortunately, access to these datasets for the average geoscience user is difficult because of the massive volumes of data generated by LiDAR. The distribution and processing (DEM generation) of large LiDAR datasets, which frequently exceed a billion data-points, challenge internet-based data distribution systems and readily available desktop software.

We are using a geoinformatics approach to the distribution and processing of LiDAR data that capitalizes on cyberinfrastructure developed by the GEON project (<http://www.geongrid.org>). Our approach utilizes a comprehensive workflow-based solution, the GEON LiDAR Workflow (GLW), which begins with user-defined selection of a subset of point data and ends with download and visualization of DEMs and derived

products. In this workflow, users perform point cloud data selection, interactive DEM generation and analysis, and product visualization all from an internet-based portal. This approach allows users to carry out computationally intensive LiDAR data processing without having appropriate resources locally.

Ultimately, we believe the GLW could be adopted as a valuable infrastructure resource for democratizing access to current and future LiDAR point cloud datasets for the geoscience community. Furthermore, the interdisciplinary geoinformatics approach taken to develop the GLW represents an excellent model for the utilization of cyberinfrastructure and information technology to tackle the data access and processing challenges presented by the next generation of geoscience data.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Ramón Arrowsmith, for his support and guidance during my MS thesis work. Ramón's expertise, enthusiasm, and energy have made him an inspiration. I am especially thankful for Ramón's encouragement and support of various opportunities for international travel and meeting presentations that have contributed significantly to my MS experience. I would also like to thank my current and former committee members, Dr. Stephen Reynolds, Dr. Mark Schmeeckle and Dr. Dan Sarewitz for their comments and recommendations on my research. Because of the truly interdisciplinary nature of the research presented here, this thesis would not have been possible without the close collaboration of my computer science colleagues Jeffrey Conner (Arizona State University), Gilead Wurman (now at UC Berkeley), Efrat Frank, Ashraf Memon, Viswanath Nandigam, and Chaitan Baru (all at San Diego Supercomputer Center). In addition to our research collaborations, these colleagues have taught me a great deal about many aspects of their field. Dr. Helena Mitsova (North Carolina State University) has provided invaluable support for questions related to the implementation of GRASS GIS in the GEON LiDAR Workflow. Her prompt and thorough responses to questions have been amazingly helpful. Dr. Carol Prentice of the U.S. Geological Survey was one of my early geology mentors and contributed to the work presented in Appendix II. Dr. Thad Wasklewicz, of the University of Memphis generously shared a portion of his Death Valley ALSM dataset for analysis in Appendix II. Financial support for my thesis research was provided by National Science Foundation grant no. 0225543 (GEON), National Earthquake Hazards Reduction Program (NEHRP) grant no. 06HQGR0032, a

grant from the Geological Society of America (GSA) Graduate Research Fund and a conference travel grant from the Arizona State University Graduate and Professional Student Association (GPSA). I am grateful to my fellow graduate students in the Active Tectonics Research Group for their camaraderie and sense of community. Finally I acknowledge my girlfriend and family for their support and encouragement during the past three years.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	ix
CHAPTER	
1 INTRODUCTION: ACTIVE TECTONICS, DIGITAL TOPOGRAPHY AND GEOINFORMATICS.....	1
Tectonic Geomorphology.....	2
Digital Topography.....	4
Geoinformatics.....	6
References Cited	9
2 A GEOINFORMATICS APPROACH TO LIDAR DATA DISTRIBUTION AND PROCESSING.....	14
Abstract.....	15
Introduction.....	17
Introduction to aerial LiDAR data: Opportunities and Challenges	19
The Computational Challenge	24
A Geoinformatics Approach.....	25
The Vision: A Conceptual Workflow for LiDAR Data Distribtuion and Processing.....	27
Proof on Concept Implementation	30
Future Work	35
Conclusions.....	37
References Cited	38

APPENDIX

I	EXPLORATION OF LIDAR POINT CLOUD DATA ARTIFACTS, RETURN DENSITY AND GENERATION OF DIGITAL ELEVATION MODELS....	61
	Summary	62
	References Cited	64
II	EXPLOITING LIDAR DATA FOR REGIONAL MORPHOLOGIC CORRELATION AND DATING OF WAVE-CUT AND FAULT-CONTROLLED LANDFORMS	80
	Summary	71
	References Cited	108

LIST OF FIGURES

Figure	Page
1.1 Comparison of different resolution topographic dataset.....	13
2.1 3D view of LiDAR point cloud	44
2.2 3D view of classified LiDAR point cloud	45
2.3 Generalized LiDAR acquisition and processing workflow	47
2.4 Full feature and bare earth LiDAR DEMs.....	49
2.5 Map showing the extent of NSAF LiDAR dataset	51
2.6 GEON Conceptual LiDAR Workflow.....	52
2.7 GEON LiDAR Workflow Portal	54-56
2.8 GEON LiDAR Workflow Products.....	57
2.9 GEON LiDAR Workflow Implementation Overview.....	58
2.10 GEON-based model for community LiDAR datasets	60
I.1 3D visualization of classified LiDAR point cloud data.....	66
I.2 LiDAR return density evaluation	68-69
I.3 Testing DEM resolution based on LiDAR ground return density	71-72
I.4 Comparison of common DEM generation algorithms	74-75
I.5 DEM generation in areas of low ground return density	78-79
II.1 Simple scarp diffusion: finite slope initial form.....	85
II.2 Illustration of morphologic dating method on synthetic “LiDAR data”..	87-88
II.3 Hanks et al., 1984 revisited – Santa Cruz, CA marine terraces	90-92
II.4 Hanks and Wallace, 1984 revisited – Lake Lahontan shoreline scarps ...	94-96

II.5 Application of morphologic dating to LiDAR/ALSM data – Sheep Creek fan
fault scarps, Death Valley, CA 99-102

II.6 Application of morphologic dating to LiDAR/ALSM data – Marine terraces,
Marine County, CA..... 105-107

CHAPTER 1

INTRODUCITON: TECTONIC GEOMORPHOLOGY, DIGITAL TOPOGRAPHY

AND GEOINFORMATICS

TECTONIC GEOMORPHOLOGY

The geoscience sub-discipline of Tectonic Geomorphology is defined as the study of landforms and landscapes that record a measurable tectonic signal (Burbank and Anderson, 2001). The overarching goal of tectonic geomorphology is to utilize the tectonic signal recorded in the landscape to understand deformation over centennial (interseismic) to millennial (Quaternary) time scales. Typically, the aim is to understand how coseismic (and aseismic) deformation is compounded to manifest itself as landforms that reflect thousands of seismic cycles – in other words, how does a mountain range thousands of meters tall grow from repeated earthquakes with 10s of meters of surface displacement? By looking at the tectonic signal recorded in landforms of varying ages, and therefore representing various stages of development, tectonic geomorphology provides a mechanism to approach this fundamental question. In addition to simply bridging the gap between time scales of landscape deformation, tectonic geomorphology also may provide insight into variations in rates of tectonic deformation over different lengths of time.

Tectonic geomorphic studies can take on a range of spatial apertures depending upon the deformation time scale of interest. Fault zone geomorphology and paleoseismic studies are interested in discerning the tectonic deformation recorded in the landscape from the past handful of earthquake cycles (e.g. Sieh, 1978). Larger spatial aperture studies such as that of the classic Wheeler Ridge along the south-central San Andreas fault (e.g. Keller et al., 1995) seek to understand the tectonic signal as recorded in a single, relatively large landform. Finally, large scale studies may aim to understand the

relationship between surface and tectonic processes at the watershed or orogen scale (e.g. Burgmann et al., 1994; Kirby et al., 2003).

Because the earth's topography is a measure of the combined effects of tectonic and surface properties (Arrowsmith, 2006), understanding the tectonic signal recorded in the landscape also requires an understanding of the surface processes that act on that landscape. These surface processes are typically a function of climate, geologic materials (bedrock/substrate), and vegetation, animal or anthropogenic factors. Specifically, processes such as linear-slope dependent transport, non-linear transport, soil production from bedrock, river incision into bedrock, landsliding, debris flows, and surface wash (Dietrich et al., 2003) among others, may act to generate or modify landforms.

Therefore, to fully understand the formation of landforms in tectonically active regions, it is necessary to address both the tectonic and surface processes at work.

Tectonic geomorphology studies traditionally employ a number of tools to understand both the tectonic and surface processes acting to shape a landform or landscape,. These tools include geologic mapping to constrain bedrock lithology, soil analyses, channel and hillslope profiles gathered via total station survey and geomorphic mapping on stereo aerial photography. More recently, tools such as remotely sensed imagery, surface exposure age dating and digital topography have supplemented the tectonic geomorphologist's tool kit.

DIGITAL TOPOGRAPHY

The availability of digital topography, in the form of digital elevation models (DEMs), has dramatically enhanced the types of analyses that are possible in tectonic geomorphic studies and as a result, the analysis of DEMs has become ubiquitous (Pike, 2002). These data enable quick calculation of common geomorphic metrics such as slope and aspect as well as more sophisticated analysis such as local relief (e.g. Burbank and Anderson, 2001) and drainage area with respect to gradient calculations (e.g. Roering et al., 1999). Digital topography also is a powerful tool for profile-based analysis of landscapes and landforms. Because profiles can be easily extracted from DEMs, users are able to efficiently acquire numerous thalweg or hillslope profiles that can then be analyzed in a variety of manners (e.g. Kirby and Whipple, 2001; Crosby et al., 2004 and Appendix II this volume). Finally, DEMs can be used as input for numerical models to determine the role of tectonic deformation in the growth of a given landform or landscape. For example, Arrowsmith (2006) uses an elastic dislocation models (e.g. Okada, 1985) to calculate the slip and uplift from a single earthquake event on the blind fault beneath Wheeler Ridge. This model predicts the general geometry of deformation associated with this single earthquake event and can be used to estimate how repeated earthquakes on this fault contribute to the growth and evolution of Wheeler Ridge.

The availability of topography in a digital form also enables these data to be combined with other digital datasets in a Geographic Information System (GIS) environment to perform analysis. The GIS environment allows the integration of topographic data with geologic and quaternary mapping, land cover, and raster imagery.

This type of data integration is also very powerful for 3D visualization of landscapes for both research and educational applications.

Digital Elevation Models are available at a variety of scales, ranging from the one kilometer resolution global coverage GTOPO30 (<http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>) dataset provided by the U.S. Geological survey to ultra-high resolution, sub-centimeter data produced from terrestrial laser scanning (TLS a.k.a. Ground-based LiDAR). Common DEM datasets utilized for tectonic geomorphic studies include the nearly global 90 m Shuttle Radar Topography Mission (SRTM) dataset (<http://srtm.usgs.gov/>) and the U.S. Geological Survey's National Elevation Dataset (NED) (<http://ned.usgs.gov/>) which includes 30 m and 10 m coverage for the United States. More recently, the rapid growth of aerial LiDAR (Light Distance And Ranging, a.k.a. ALSM (Airborne Laser Swath Mapping)) for earth science applications has provided high-resolution topographic data (sub-meter to 5 m resolution) across large swaths of the United States (see the Chapter 2 section entitled "Introduction to Aerial LiDAR Data: Opportunities and Challenges" in this volume for a full introduction to LiDAR data).

LiDAR data is quickly becoming one of the most powerful tools in the earth sciences for studying the earth's surface. Capable of generating digital elevation models (DEMs) more than an order of magnitude more accurate than those currently available via the USGS National Elevation Dataset, LiDAR data offer geomorphologists the opportunity to study natural processes at resolutions not previously possible yet essential for their appropriate representation. Because of the high resolution of LiDAR-derived

DEMs, subtle geomorphic features that would be lost in coarser 10 and 30 m terrain models are captured and can thus be quantified as components of the geomorphic process acting to shape a given landscape or landform (Figure 1.1).

However, as I discuss in the Chapter 2 section entitled “The Computational Challenge”, the distribution and processing of LiDAR data for geoscience users presents a significant challenge. In order to make these powerful yet computationally challenging datasets useful for the tectonic geomorphology and greater geoscience communities, a new approach to their management, archiving, distribution, processing and integration is necessary.

GEOINFORMATICS

The emerging science discipline of Geoinformatics (e.g., Sinha, 2000) seeks to build a shared cyberinfrastructure for the geosciences through interdisciplinary collaboration between earth and computer scientists. The goal of this geoscience cyberinfrastructure is to:

...(1) manage, preserve, and efficiently access the vast amounts of Earth science data that exist now and the vast data flows that will be coming online as projects such as EarthScope get going; (2) foster integrated scientific studies that are required to address the increasingly complex scientific problems that face our scientific community; (3) accelerate the pace of scientific discovery and facilitate innovation; (4) create an environment in which data and software developed with public funds are

preserved and made available in a timely fashion; and (5) provide easy access to high-end computational power, visualization, and open-source software to researchers and students (Owens and Keller, 2003).

Because of the massive volumes of data and the computational challenges they present, aerial LiDAR data is an excellent example of a dataset that would benefit from a geoinformatics approach to its management, archiving, distribution, processing and integration.

In Chapter 2 of this thesis, I present a geoinformatics-based approach to LiDAR data distribution and processing – the GEON LiDAR Workflow. The fundamental goal of that project is to democratize access to aerial LiDAR point cloud data for the geoscience community. By allowing users to perform point cloud data selection, interactive DEM generation and analysis, and product visualization all from an internet-based portal, we hope to promote the use of high-resolution topographic data in geoscience research. Clearly, the current implementation of the GLW discussed in Chapter 2 is of most interest to tectonic geomorphology and earth surface processes researchers. However, we believe that the GLW could be adopted as a valuable infrastructure resource for democratizing access to current and future LiDAR point cloud datasets for the whole geoscience community. Because these datasets also capture vegetation and built structures in addition to topography, we expect that researchers in the ecological, engineering and urban planning communities could also capitalize upon the data and tools available through the GLW.

Although aerial LiDAR data and the GEON approach to these datasets is an excellent demonstration of the development of a cyberinfrastructure-based toolset, it is but one example of how geoinformatics is being applied to develop community-oriented and internet-based databases, tools and computing resources. For example, the Southern California Earthquake Center's (SCEC) Information and Technology Research (ITR) project seeks to use cyberinfrastructure to develop a "Community Modeling Environment (CME)" (<http://epicenter.usc.edu/cmeportal/index.html>) oriented towards system-level earthquake science (Jordan et al., 2003). This "rupture to rafters collaboratory" was conceived to provide an environment where basic geophysical research (e.g Zhao et al., 2004) can be combined with seismic hazard modeling (Field et al., 2003) and earthquake engineering in an environment where tools, databases and computing resources are integrated by Information Technology. Yet another example of the informatics approach to community geoscience cyberinfrastructure is the NEPTUNE Project (<http://www.neptune.washington.edu/index.jsp>) which plans to deploy a regional cabled ocean observatory on the Juan de Fuca plate off the coast of the Pacific Northwest. In the NEPTUNE project, cyberinfrastructure and will be used to monitor and control the observation network, distribute raw and processed data products, and to "provide the computer-based toolsets needed by scientists, engineers, and others to collect and process information generated by the system" (<http://www.neptune.washington.edu/infrastructure/index.jsp?keywords=NETWRK&title=Network%20Management>)

Although the GLW, SCEC CME and NEPTUNE projects were only briefly summarized above, I think they are excellent examples of the growing acceptance that geoinformatics provides a powerful new approach for managing geoscience data and tools. As geoscience data acquisition technologies continue to become more sophisticated and their data streams more massive, the development of community oriented toolsets that provide access to data as well as processing tools and computing resources are going to become increasingly necessary. In many respects, I believe that the geoinformatics approach presented in this thesis for LiDAR point cloud data distribution and processing represents the future of resources designed to facilitate access and processing for the next generation of earth science data.

REFERENCES CITED

- Arrowsmith, J R., 2006, Active tectonics, tectonic geomorphology, and fault system dynamics: how geoinformatics can help, in Sinha, A. K., ed., *Geoinformatics: Data to Knowledge: Geological Society of America Special Paper 397*, p. 131-139.
- Burbank, D.W. and Anderson, R.S., 2001, *Tectonic Geomorphology*: Malden, Massachusetts, Blackwell Science, 274 p.
- Burgmann, R., Arrowsmith, R., Dumitru, T. A., and McLaughlin, R. J., 1994, Rise and fall of the southern Santa Cruz Mountains, California, from Fission tracks, geomorphology, and geodesy: *Journal of Geophysical Research*, 99, 20,181-20,202.

Crosby, C. J., Arrowsmith, J R., Oldow, J. S., Prentice, C. S., 2004, Exploiting LiDAR for Regional Morphologic Correlation and Dating of Wave-cut and Fault-Controlled Landforms: *Eos Trans. AGU*, 85 (47), Fall Meet Suppl., Abstract G13B-0810.

Field, E.H., Jordan, T.H, and Cornell, C.A., 2003, OpenSHA: A Developing Community-modeling Environment for Seismic Hazard Analysis: *Seism. Res. Lett.*, v. 74, p. 406-419.

Jordan, T.H., Maechling, P.J., and SCEC/CME Collaboration, 2003, The SCEC Community Modeling Environment – An Information Infrastructure for System-Level Earthquake Science: *Seism. Res. Lett.*, v. 74, no. 3, p. 44-46.

Keller, E. A., Zepeda, R. L., Rockwell, T. K., Ku, T. L. and Dinklage, W. S., 1998, Active tectonics at Wheeler Ridge, southern San Joaquin Valley, California: *Geological Society of America Bulletin*, v. 110; no. 3; p. 298-310.

Kirby, E., Whipple, K.X., Tang, W., and Chen, Z., 2003, Distribution of active rock uplift along the eastern margin of the Tibetan Plateau: Inferences from bedrock river profiles: *Journal of Geophysical Research*, v. 108, doi: 10.1029/2001JB000861.

Kirby, E., and Whipple, K., 2001, Quantifying differential rock-uplift rates via stream profile analysis: *Geology*, v. 29, p. 415-418.

Lawson, A.C. (ed.), 1908, *The California Earthquake of April 18, 1906*: Reprinted 1969 by the Carnegie Institution of Washington, D.C.

Owens, T.J. and Keller, G.R., 2003, GEON (GEOscience Network): A First Step in

Creating Cyberinfrastructure for the Geosciences: *Electronic Seismologist*, v. 74,
n. 4

Pike, R.J., 2002, A bibliography of terrain modeling (geomorphometry), the quantitative representation of topography; supplement 4.0: U.S. Geological Survey Open File Report 02-465.

Sieh, K.E., 1978, Slip along the San Andreas fault associated with the great 1857 earthquake: *Bulletin of the Seismological Society of America*; October 1978; v. 68; no. 5; p. 1421-1448.

Zhao, L., Chen, P., Jordan, T H., Olsen, K B., Maechling, P., Faerman, M., 2004, Cyberinfrastructure for the Unified Study of Earth Structure and Earthquake Sources in Complex Geologic Environments: *Eos Trans. AGU*, 85(47), Fall Meet. Suppl., Abstract SF31B-02.

Figure 1.1. A) USGS Digital Orthophoto Quarter Quadrangle (DOQQ) for Mill Gulch near Fort Ross, CA along the trace of the northern San Andreas fault (NSAF) (southern-most portion of area shown in Figure 2.5). This grey scale raster image has a pixel resolution of 1 meter. Note the representation of geomorphic features such as the right-laterally offset drainage due to slip on the NSAF, active landslides, vegetation, and human influences such as buildings and roads. B) Hillshade of a USGS 30 m DEM for the same area as shown in (A). The coarseness of this terrain model obscures many of the small, yet important geomorphic features of the landscape that are visible in the orthophoto in (A). C) Hillshade of a full feature LiDAR derived DEM produced using the GEON LiDAR Workflow (Chapter 2, this volume). In this 1.8 m DEM, many of the geomorphic features visible in the orthophoto (A) are captured in the terrain model. In addition, the DEM provides 2.5D representation of the landscape that can be used for visualization and modeling. D) 5 ft contour interval map of the Mill Gulch stretch of the 1906 Earthquake rupture made by Francois Matthes in the months following the earthquake (Lawson, 1908). The Matthes map, produced with plane table and alidade, demonstrates that high-resolution topography has always been of interest for scientists studying the expression of tectonic deformation on the landscape.

