

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:

Director of the School

Dean, Division of Graduate Studies

ABSTRACT

The emergence of new digital data acquisition technologies in the geosciences has important implications for the types and quality 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 use 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, 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 | 1 |
| Digital Topography..... | 3 |
| Geoinformatics..... | 5 |
| References Cited | 8 |
| 2 A GEOINFORMATICS APPROACH TO LIDAR DATA DISTRIBUTION AND PROCESSING FOR THE EARTH SCIENCES..... | 13 |
| Abstract..... | 13 |
| Introduction..... | 15 |
| Introduction to aerial LiDAR data: Opportunities and Challenges | 17 |
| The Computational Challenge | 22 |
| A Geoinformatics Approach..... | 24 |
| The Vision: A Conceptual Workflow for LiDAR Data Distribtuion and Processing..... | 25 |
| Proof on Concept Implementation | 28 |
| Future Work | 33 |
| Conclusions..... | 35 |
| References Cited | 37 |

APPENDIX

| | | |
|----|--|-----|
| I | EXPLORATION OF LIDAR POINT CLOUD DATA ARTIFACTS, RETURN DENSITY AND GENERATION OF DIGITAL ELEVATION MODELS.... | 59 |
| | Summary | 59 |
| | References Cited | 61 |
| II | EXPLOITING LIDAR DATA FOR REGIONAL MORPHOLOGIC CORRELATION AND DATING OF WAVE-CUT AND FAULT-CONTROLLED LANDFORMS | 77 |
| | Summary | 77 |
| | References Cited | 104 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 1.1 Comparison of different resolution topographic dataset..... | 11 |
| 2.1 3D view of LiDAR point cloud | 42 |
| 2.2 3D view of classified LiDAR point cloud | 43 |
| 2.3 Generalized LiDAR acquisition and processing workflow | 44 |
| 2.4 Full feature and bare earth LiDAR DEMs..... | 46 |
| 2.5 Map showing the extent of NSAF LiDAR dataset | 48 |
| 2.6 GEON Conceptual LiDAR Workflow..... | 50 |
| 2.7 GEON LiDAR Workflow Portal | 51 |
| 2.8 GEON LiDAR Workflow Products..... | 55 |
| 2.9 GEON LiDAR Workflow Implementation Overview..... | 56 |
| 2.10 GEON-based model for community LiDAR datasets | 57 |
| I.1 3D visualization of classified LiDAR point cloud data..... | 62 |
| I.2 LiDAR return density evaluation | 64 |
| I.3 Testing DEM resolution based on LiDAR ground return density | 67 |
| I.4 Comparison of common DEM generation algorithms | 70 |
| I.5 DEM generation in areas of low ground return density | 73 |
| II.1 Simple scarp diffusion: finite slope initial form..... | 80 |
| II.2 Illustration of morphologic dating method on synthetic “LiDAR data” | 82 |
| II.3 Hanks et al., 1984 revisited – Santa Cruz, CA marine terraces | 85 |
| II.4 Hanks and Wallace, 1984 revisited – Lake Lahontan shoreline scarps | 89 |

| | |
|---|----|
| II.5 Application of morphologic dating to LiDAR/ALSM data – Sheep Creek fan fault scarps, Death Valley, CA | 93 |
| II.6 Application of morphologic dating to LiDAR/ALSM data – Marine terraces, Marine County, CA..... | 99 |

CHAPTER 1

INTRODUCTION: 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 in landforms that reflect thousands of seismic cycles – in other words, how does a mountain range thousands of meters tall grow from repeated earthquakes with 1-10s 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; Burbank and Anderson, 2001; 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 (in the US, 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, 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

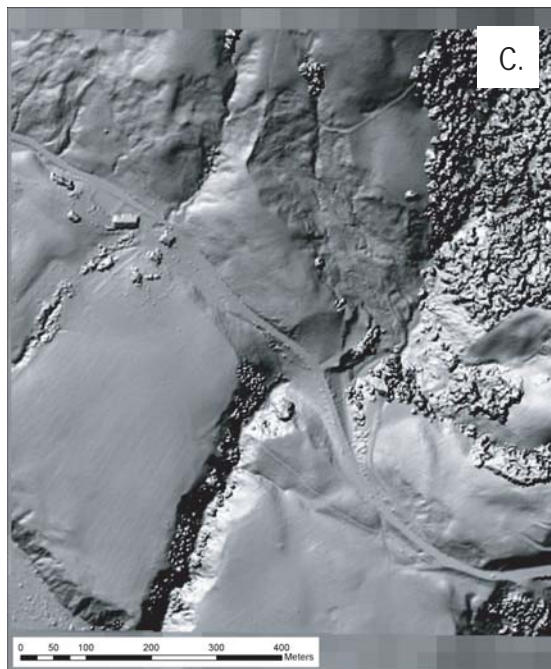
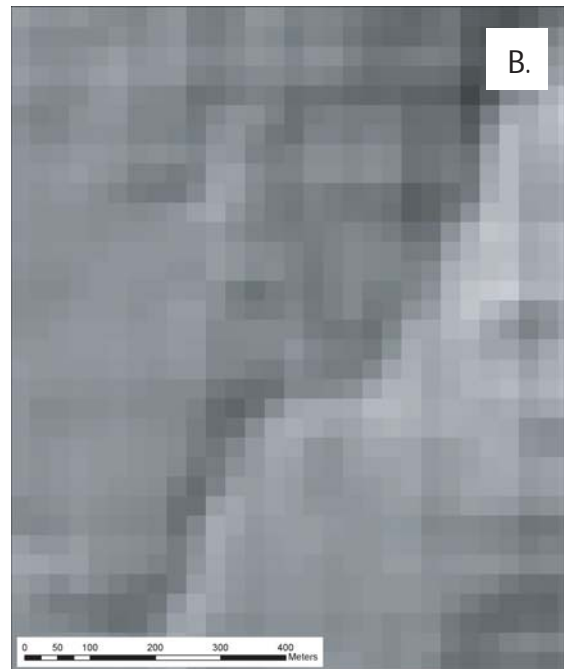
Creating Cyberinfrastructure for the Geosciences: *Electronic Seismologist*, v. 74,
n. 4

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.



CHAPTER 2

A GEOINFORMATICS APPROACH TO LIDAR DATA DISTRIBUTION AND PROCESSING FOR THE EARTH SCIENCES

ABSTRACT

The emergence of new digital data acquisition technologies in the earth sciences has important implications for the types and quality of data that are now available to researchers. However, along with these new 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 earth science applications are an excellent example of the opportunities and challenges presented by these types of datasets. LiDAR data is quickly becoming one of the most powerful tools in the earth sciences for studying the earth's surface and overlying vegetation. Capable of generating digital elevation models (DEMs) more than an order of magnitude more accurate than those currently available, LiDAR data offer scientists the opportunity to study natural processes at resolutions and extents not previously possible yet essential for their appropriate representation.

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

users who lack the computing and software resources necessary to handle these enormous and high-density datasets.

We use a geoinformatics approach to the distribution and processing of LiDAR data that capitalizes on cyberinfrastructure being developed as part of the GEON project. Our approach utilizes a comprehensive workflow-based solution, the GEON LiDAR Workflow (GLW), which begins with user-defined selection of a subset of raw point data and ends with download and visualization of digital elevation models and derived products. In this workflow, the billions of points within a LiDAR dataset point cloud are hosted in a spatially indexed database. Data selection and processing is performed via an internet portal that allows users to execute spatial and attribute subset queries on the larger dataset. The subset of data is then passed to a web service built with Open Source software for interpolation to grid and analysis of the data. The interpolation and analysis portion of the workflow offers spline interpolation to grid with user-defined grid (DEM) resolution as well as control over the spline parameters. We also compute geomorphic metrics such as slope, curvature, and aspect as derived products from the DEM that the GLW generates. Users may choose to download their results in ESRI or ASCII grid formats as well as geo tiff.

Ultimately, the GEON LiDAR Workflow 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 GEON LiDAR Workflow 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 earth science data.

INTRODUCTION

Recent advances in digital data acquisition technologies in the earth sciences are having a profound impact on the types and quality of datasets that are available to researchers. Technologies such as real-time sensor networks, space and airborne-based remotely sensed data, real-time geodetic and seismologic data, massive geospatial databases and the outputs of large computational model runs are all enabling new and exciting research on the forefront of the earth sciences. However, with these technologies comes a prodigious 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. Unfortunately, data processing expertise or computing resource requirements may sometimes be a barrier to accessing these massive datasets for the broader scientific community. An emerging solution for enhancing access to these datasets is to develop a shared cyberinfrastructure that uses information technology to provide access to data, tools and computing resources. Owens and Keller (2003) describe cyberinfrastructure for the earth sciences as:

...the information technology infrastructure that is needed 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.

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 geoinformatics approach of harnessing the tools of information technology to facilitate earth science research provides a mechanism to equalize and simplify access to the next generation of earth science data.

Aerial LiDAR (Light Distance And Ranging), also known as ALSM (Airborne Laser Swath Mapping), data have recently emerged as one of the most powerful tools available for documenting the earth's topography and its masking vegetation at very high-resolution. LiDAR derived digital elevation models (DEMs) are typically more than an order of magnitude more accurate than the best-available U.S. Geological Survey 10 meter DEMs. The ability to use these data to construct two-and-a-half and three-dimensional models of the earth's topography and vegetation are rapidly making them an indispensable tool for earth science research (Carter et al., 2001). Unfortunately, due to the massive volumes of data associated with this technology, accessing and processing these datasets can be challenging for even the most sophisticated LiDAR users.

In this paper we present a geoinformatics approach to the internet-based distribution and processing of LiDAR data, with particular emphasis on research-grade, community datasets. This approach could be adopted as a model for the utilization of cyberinfrastructure and information technology to tackle the data access and processing challenges presented by the next generation of earth science data. The development of community oriented data portals as well as internet-based processing, mapping and visualization tools is applicable to many types of earth science data and is likely to revolutionize the utilization of large and complex datasets.

This paper provides an introduction and overview of LiDAR technology and its applications, followed by a discussion of the computationally challenging aspects of these datasets. Next, we propose a conceptual, work-flow based approach to the internet-based distribution and processing of these data. Finally, we present a proof of concept implementation of the workflow and discuss future developments for this project.

INTRODUCTION TO AERIAL LIDAR DATA: OPPORTUNITIES AND CHALLENGES

Acquisition of aerial LiDAR data utilizes a pulsed laser ranging system operating at 10s of thousands of pulses per second mounted in an aircraft equipped with a kinematic Global Positioning System (GPS) to provide precise positioning information for the aircraft. An accurate inertial measurement unit (IMU) monitors the orientation (roll, yaw and pitch) of the aircraft. By combining the laser ranging system, GPS, and IMU on an aircraft, it is possible to quickly and economically acquire hundreds of

millions to billions of individual point measurements of the absolute x, y and z coordinates of the ground surface and vegetative cover in a survey area. LiDAR instruments typically sample the ground surface multiple times per square meter and provide an absolute vertical accuracy of 5-10 cm (Shrestha et al., 1999; Wehr and Lohr, 1999). These LiDAR measurements are commonly referred to as the 'point cloud' (x,y,z plus attributes) and typically consist of 100s of millions or billions of returns depending upon the size of the survey area and resolution of the data being acquired (Figure 2.1). Because most modern LiDAR instruments are capable of recording multiple returns from each outgoing laser pulse, it is possible to classify the individual laser returns by applying a filtering algorithm to differentiate ground returns from vegetation returns (e.g. Haugerud and Harding, 2001; Sithole and Vosselman, 2004) (Figure 2.2). This ability to segregate the point cloud data based upon the origin of the return significantly enhances the utility of these datasets to a wide variety of user communities.

A generalized aerial LiDAR acquisition and processing workflow consists of the following four steps: 1) Data acquisition, 2) processing of laser ranging, GPS and IMU data to generate LiDAR point cloud, 3) point cloud classification and 4) generation, manipulation, and delivery of digital ground and vegetation models (Figure 2.3).

Typically the first half of this workflow, data acquisition and point cloud generation (and sometimes the point cloud classification), is handled by the data provider and is not a significant concern to the researcher seeking LiDAR products for science applications. The later portion of this workflow however should be of great interest to scientific users because point cloud classification and generation of digital ground and vegetation models

(Figure 2.4) directly controls the character of the products upon which analyses will be performed. A small percentage of earth science users may have interest in the data acquisition and point cloud generation steps in the workflow for geodetically oriented studies (e.g. Bevis et al., 2005). Therefore, for most scientific applications it is appropriate to consider the LiDAR point cloud the ‘raw’ data product.

Because the management and processing (e.g., filtering, digital elevation model generation, and manipulation) of LiDAR point cloud data requires sophisticated computing resources and knowledge [and because their science application requires it], the majority of earth science LiDAR users typically perform their analyses on ‘bare-earth’ digital elevation models (DEMs) derived from LiDAR ground returns. These digital files are created by interpolating or binning the LiDAR point cloud to a continuous surface and then converting this surface to a regularized grid with elevation data sampled at uniformly spaced intervals (e.g. Mitas and Mitasova, 1999). Because of the computational difficulty of DEM generation, LiDAR DEMs - often bare-earth (ground only) and full-feature (unfiltered) - are generated once by the data provider and then the raw point cloud data is essentially discarded. In some cases the user may not even take delivery of the point cloud data. Unfortunately, users who work only with the standard vendor generated DEM products are not to take full advantage of the richness of information held in the point cloud.

By beginning analyses with the raw point cloud data, users gain an understanding of the data and control over how those data are used to characterize the landscape and vegetative cover. Details such as the interpolation algorithm and grid resolution can

significantly affect the manner in which the resulting DEM represents the landscape. In addition, beginning with the LiDAR point cloud data allows the user to assess the homogeneity and density of ground returns, which can vary due to topography, canopy characteristics and acquisition parameters, and to evaluate potential artifacts in the data caused by errors in data acquisition and processing (point misclassification) (Appendix I) (Crosby and Arrowsmith, 2006).

The availability of digital topographic data of the resolution and accuracy provided by aerial LiDAR has had profound implications for research in earth surface processes, natural hazards, ecology and engineering (Carter et al., 2001; Stoker et al., 2006). For example, in the geosciences, aerial LiDAR data has been used to document ground rupture following the 1999 Hector Mine earthquake (Hudnut et al., 2002), measure topographic change during the 2004 eruption of Mount St. Helens, WA (Haugerud et al., 2004), calibrate and test hillslope transport laws (e.g., Roering et al., 1999), quantify beach topography changes (e.g. Sallenger et al., 2003) and to improve flood hazard and floodplain maps under the FEMA National Flood Insurance Program (e.g. North Carolina Floodplain Mapping Program, 2000). In addition to the wide spectrum of geoscience applications for LiDAR data, the ecology community also uses aerial LiDAR data for its ability to provide three-dimensional measurements of vegetation canopies and to examine the correlation between topography and vegetation. Lefsky et al. (2002) provide an overview of applications of LiDAR data to measuring vegetation height, biomass and canopy structure and function in the context of ecosystem studies. Finally, the high-resolution topography provided by aerial LiDAR data is being

utilized in urban environments for urban planning, built-structure feature extraction and three-dimensional city visualization (e.g. Haala and Brenner, 1999; Maas and Vosselman, 1999).

As these examples illustrate, there is rapidly expanding interest in, and utilization of, aerial LiDAR data in scientific research. Due to this increased demand, LiDAR data acquisition has exploded in the past 5 to 10 years with more than 120 LiDAR acquisition systems in use worldwide (Espinosa et al., 2006). Although it is difficult to count the number of publicly available LiDAR datasets due to a heterogeneous network of aerial LiDAR data providers and funding agencies, it is reasonable to assume there are 100s of billions of LiDAR point returns in the public domain and available for earth science research. Many of these data were gathered with a single scientific or management purpose, yet they actually are useful as community datasets with numerous applications that can be enabled if they are made appropriately available. With typical LiDAR systems operating at 10s of kilohertz and cutting edge systems operating at 85-100 kilohertz (Baltasvias, 1999), it is increasingly common for a single LiDAR point cloud dataset to contain 100s of millions or billions of points (Figure 2.5). As LiDAR systems continue to mature, the size of these datasets will only continue to expand. In addition, the simultaneous acquisition of raster imagery (aerial and hyperspectral) with LiDAR data is becoming increasingly common. These raster datasets should be hosted together with the LiDAR and add significantly to the total size of the dataset. Although this vast and rapidly expanding volume of aerial LiDAR data represents a significant opportunity for new scientific endeavors, it also presents a massive challenge in terms of

management, distribution and processing. The current demand for, and interest in, aerial LiDAR data far outpaces the resources available within the earth science community for the distribution and processing of these data.

THE COMPUTATIONAL CHALLENGE

Public domain LiDAR point cloud data are typically delivered by the data provider to the funding agency or primary investigator in a generic ASCII format or in the binary LAS format (ASPRS, 2005) on a portable hard drive or DVDs. The first challenge arises in the act of distributing these massive public domain datasets to the scientific community. Mailing DVDs or hard drives to users who request the data is an inefficient and time consuming solution. A more elegant approach to the problem is to utilize high bandwidth cyberinfrastructure and information technology to distribute the data via an interactive, internet-based, portal. The National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center's LIDAR Data Retrieval Tool (LDART) (<http://www.csc.noaa.gov/crs/tcm/about/ldart.html>) and the U.S. Geological Survey's Center for LIDAR Information Coordination and Knowledge (CLICK) (Stoker et al., 2006) are good examples of ongoing efforts to provide internet-based access to LiDAR point cloud data. The NOAA system is devoted to NOAA coastal LiDAR data and is therefore limited in its scope. The USGS CLICK effort is much broader as it seeks to be an archive and distribution resource for raw, unfiltered, LiDAR point cloud data from across the United States.

Although a centralized resource for locating and accessing LiDAR point cloud data such as CLICK is an excellent first step towards allowing earth scientists to utilize raw LiDAR data in their research, it does not fully address the challenges associated with LiDAR data distribution and processing. The CLICK data distribution system hosts and serves LiDAR point cloud data organized as USGS quarter quadrangles in either generic ASCII or LAS binary formats. Because of the massive number of returns in a single LiDAR dataset, a single quarter quadrangle's worth of data may still contain 10s of millions to 100s of millions of returns and result in file sizes of 100s of megabytes (in the smaller LAS binary format). Additionally, by serving data only in quarter quadrangles, users are forced to download and merge multiple data tiles if they are interested in a region that falls on a quadrangle boundary.

Although nearly ubiquitous high-speed internet access now makes it possible to download massive LiDAR point cloud data files such as those provided by CLICK, visualization and DEM generation for these data typically require complex and typically expensive software packages (see U.S. Army Topographic Engineering Center, 2006 for software examples) that may, for financial, computing resource or expertise reasons, be beyond the reach of the average earth scientist. This computational barrier to working with LiDAR data is therefore the current limiting factor for earth science users who wish to incorporate LiDAR point cloud data analysis and processing into their research.

A GEOINFORMATICS APPROACH

Because of the data volumes and processing difficulty associated with LiDAR point cloud data, it is unrealistic to expect all earth science users who wish to work with these datasets to independently acquire the tools and resources necessary. An enticing alternative is to employ an interdisciplinary, geoinformatics approach that utilizes cyberinfrastructure resources to build a community-oriented data distribution and processing toolset that all LiDAR users can use. The National Science Foundation large Information and Technology Research program-funded Geoscience Network (GEON) project (<http://geongrid.org/>) provides the resources and expertise to undertake the development of such a resource (Owens and Keller, 2003).

GEON was designed as an equal collaboration between Information Technology (IT) and geoscience researchers, with the goal of developing an enabling IT platform to facilitate the next generation of geoscience research and education. Project participants and partners include a number of US universities, federal agencies, and industry, as well as international partners. The core cyberinfrastructure that is being developed is broadly applicable beyond the geosciences to a variety of other science disciplines and application domains. GEON is based on a “service-oriented architecture (SOA)”. This architecture takes advantage of a distributed network of datasets, tools and computing resources to provide access to high performance computing platforms for data analysis and model execution. “Advanced information technologies are being developed in the GEON project to support “intelligent” data searching, semantic integration, and visualization of multidisciplinary information as well as 4D scientific datasets and

geospatial data” (<http://www.geongrid.org/about.html>). The GEON Portal provides a web-based interface and framework to access the various resources.

In addition to the computing resources provided by the project, GEON is designed to bridge cultural and disciplinary boundaries to bring together geoscientists and computer science experts for the common goal of developing the next generation of geoscience tools. This close collaboration enables GEON researchers to identify opportunities and tackle problems that they might not otherwise be equipped to handle within their own discipline. Therefore, GEON provides an exciting suite of tools (hardware) and computer science skills (people & knowledge) that can be applied to the LiDAR distribution and processing challenge faced by the earth science community.

THE VISION: A CONCEPTUAL WORKFLOW FOR LIDAR DATA DISTRIBUTION AND PROCESSING

Given the distribution and processing challenges presented by LiDAR point cloud data, we have created a conceptual workflow for an internet-based LiDAR distribution and processing toolset (Figure 2.6). This workflow capitalizes on cyberinfrastructure and information technology expertise and resources available via the GEON project to offer interactive point cloud data distribution, generation of digital elevation models, and analysis of large LiDAR datasets. The GEON conceptual LiDAR workflow is an end-to-end approach beginning with data distribution and ending with download and or visualization of products in two and three dimensions. The workflow is designed to utilize a modular, web service-based architecture that is scalable and dynamic. The goal

of the modular approach is to allow new tools and resources to be easily added and to allow the workflow to be customized on-the-fly based upon the processing selections made by the user.

In the GEON conceptual workflow, LiDAR point cloud data are hosted in a spatially indexed database which is accessed through an interactive web-portal. By storing the point cloud data in a database, users are able to perform spatial and attribute-based queries on the data. For example, the user can select all of the ground returns (attribute) within a given polygon (spatial) drawn on an interactive map in the web portal. [the query is converted from the portal interaction to an SQL statement in the background]. This approach allows users to select just the data in which they are interested and avoids the problems incurred by hosting data in tiles. We also propose hosting raster imagery that was acquired in tandem with the LiDAR data in a similar database so that these data can also be interactively accessed by users.

Once the user has selected their data of interest and the working geographic projection, sophisticated users may wish to exit the workflow by downloading the point cloud to their local computer to perform their own processing and analysis. For these users, the workflow acts only as a LiDAR point cloud data distribution resource. Other users however, may choose to take advantage of an interpolation and analysis toolset where they can perform tasks such as point cloud classification, evaluation of return density and classification errors, and DEM generation and analysis. Finally, users are able to download their data products in a variety of common file formats. By providing data products in a wide variety of formats, users are able to bring the data into their

software package of choice for additional scientific analysis. In certain situations, users may wish to view their data products before, or in place of, downloading to a local computer. In this case, the workflow offers the option of either two or three dimensional visualization of data products within the web browser. This functionality could be employed for verifying data products before beginning large downloads, for users who lack visualization and mapping resources locally or for educational purposes where students use the workflow to explore landscapes digitally.

For geoscience users, the generation of DEMs is likely the most important toolset available within the GEON conceptual LiDAR workflow because DEMs derived from LiDAR are the most frequently used LiDAR product for scientific analysis (and 2.5D visualization). Because DEM generation can be performed with a number of different interpolation and binning algorithms, the aim of the workflow is to provide the user with a suite of algorithms (i.e. spline, TIN, IDW, Kriging, block mean/min/max) along with control over algorithm parameters to allow fine-tuning of the resultant surface (Appendix I). The user is also given control over the resolution of the resultant DEM. The goal is to provide an interactive processing environment for iteration and exploration of various DEM generation algorithms and processing options. Ultimately, these tools can be used to optimize landscape representation based upon the user's scientific applications.

By utilizing the distributed computing resources available through the GEON project, the conceptual workflow allows the user to quickly and easily run multiple jobs with different algorithms and/or parameter settings and compare results. Similar iteration may take days or weeks if done locally on a single computer. The conceptual LiDAR

workflow is designed to allow computationally intensive data processing to be handled by the cyberinfrastructure available through GEON and to offer the user manageable download products in common file formats relatively quickly. This Geoinformatics approach effectively removes many of the computational barriers to working with LiDAR data and thereby democratizes access to these datasets.

PROOF-OF-CONCEPT IMPLEMENTATION

To illustrate our geoinformatics approach to aerial LiDAR point cloud data distribution and processing, we have built a proof-of-concept toolset, the GEON LiDAR Workflow (GLW), which offers a subset of the functionality discussed above (red pathways in Figure 2.6). The GLW provides users access to LiDAR point cloud data, DEM generation and analysis algorithms and download and visualization of data products all through an internet-based web portal. The GLW portal is hosted within the GEON portal (<http://portal.geongrid.org>) where users can access a variety of geoscience data and applications.

The GEON LiDAR Workflow capitalizes on a series of distributed computing resources to complete the three main processing tasks within the workflow: selection of a subset of point cloud data, generation and analysis of digital elevation models, and download and visualization of the resulting products. Coordination of these resources is handled through an architecture developed by the GEON project (Jaeger-Frank et al., 2006). This architecture uses the Kepler scientific workflow system (Ludäscher et al., 2005) to link the various databases, processing tools and computing resources. Because

each tool in the processing workflow is designed to be a modular web service, Kepler can dynamically generate a custom workflow - by linking the appropriate modules - based upon the processing parameters selected by the user at the portal. The modularity of the GLW architecture also allows us to easily add new databases, processing tools and computing resources, making the architecture adaptable and extensible.

The proof of concept GLW currently features three exemplary LiDAR point cloud datasets: 1) Data collected along the Northern San Andreas Fault (NSAF) and associated marine terraces in Sonoma and Mendocino Counties, California (Figure 2.5), 2) data from the Western Rainier Seismic Zone in Pierce County, Washington, 3) data covering active faults in the Eastern California Shear Zone portion of the Mojave Desert, California, and 4) the Dragon's Back Ridge portion of the southern San Andreas Laser Scan dataset (a.k.a "the B4 Project"). Each of these datasets is hosted in an IBM DB2 spatially indexed database running on the terascale supercomputer DataStar (http://www.sdsc.edu/user_services/datostar/) at San Diego Supercomputer Center. DataStar is optimized for data-intensive computing tasks and is therefore able to efficiently handle spatial and attribute-based queries on these massive LiDAR point cloud datasets.

The LiDAR point cloud datasets, and associated processing tools, are accessed via an internet-based portal (Figure 2.7). This portal features an interactive web map as well as check and dialog boxes which allow the user to query the dataset and control processing parameters. The interactive Web Map Service (WMS) map provides basic geospatial data layers such as roads, cities, topography and bodies of water that help the

user to understand the geographic context of the dataset. Once the user has located the data of interest, he or she draws a box on the map that defines the spatial extent of the query to be submitted to the database. In addition to the spatial query on the database, users are also able to query by the return classification attribute (vegetation, ground, structure etc. if available) so that they can more accurately select the data of greatest scientific interest (Figure 2.7B).

Once the parameters of the database query have been defined, the user can verify the number of LiDAR returns in their selection and make choices about the download products, and file formats for those products they would like to receive. The GLW currently offers generation of digital elevation models via a spline with regularized smoothing and tension (rst) algorithm (Mitasova and Mitas, 1993) available in the GRASS Open Source Geographic Information System (GIS) package (Neteler and Mitasova, 2004). At the web portal, users are given control over the DEM resolution as well as spline algorithm parameters such as tension and smoothing (Figure 2.7C). By having control over processing parameters, users are able to experiment to find the settings that produce a DEM best suited to their scientific applications. Using algorithms in GRASS GIS, the GLW also allows users to calculate the common geomorphic metrics of slope, aspect and profile curvature, as products derived from the DEM they have generated (Mitasova and Hofierka, 1993). Finally, the GLW offers users the ability to download the raw point cloud data as well as their DEM and derivative products (Figure 2.8). Point cloud downloads are offered in a compressed ASCII format while the DEMs and derivatives can be downloaded in ARC ASCII and standard ASCII grid formats or as

a GeoTIF. Upon submission of their job at the web portal, the user can choose to wait for their results to be delivered to the browser or they may close the window and an email notification will be sent when the job is complete.

Out of sight to the user, the submission of a job at the web portal executes a sequence of processing steps that are tailored to fulfill the user request. This sequence of processing steps is performed on a geographically distributed network of computing resources orchestrated by the Kepler workflow system (Figure 2.9). As such it represents a powerful example of cyberinfrastructure at work. Not only are the software and data components modular, diverse compute resources may be “plugged in” as they become necessary and/or available.

The GLW portal is served from a computer at SDSC while the LiDAR point cloud data are hosted on the separate DataStar supercomputer at SDSC (Figure 2.9). Upon submission of a job, the user’s spatial and attribute query information is sent to DataStar, where the query returns all points with the given attribute value that reside within a user selected bounding box. Next, the “clipped” subset of data, along with the processing parameters, is sent from SDSC to Arizona State University where a GEON computing cluster is waiting to perform the data processing. This transfer of data occurs within seconds and takes advantage of high-bandwidth internet connections and tools such as the Grid File Transfer Protocol (Grid FTP). A GRASS GIS web service deployed on the ASU cluster accepts the data and processing parameters and then executes the request as the user defined it at the GLW portal. The bulk of the computational run time in the workflow is spent interpolating the point cloud data to produce a DEM using the

spline algorithm within GRASS GIS. Once the DEM and derived products are complete, they are compressed and shipped back to SDSC where yet another computer generates browse images of the products using Global Mapper (<http://www.globalmapper.com/>). Finally, a web page with the browse images and links to a directory containing the user's products is produced and an email is sent to notify the user that their job is complete.

Given the current architecture and processing tools, a job submitted to the GLW will run to completion in less than 10 minutes. At present we limit jobs to a maximum of 1.6 million points for DEM generation and 20 million points for download. Interpolation of large volumes of LiDAR point cloud data currently represents a bottleneck in the overall GLW workflow. The implementation of parallelized interpolation code or alternative, more efficient, binning or interpolation algorithms are likely to result in significantly improved processing performance.

Because the GLW architecture is designed to access and utilize distributed databases and computing resources (Figure 2.9), the system can easily be adapted to expose new resources as they become available. Although the LiDAR data distributed via the GLW are currently hosted by GEON at the San Diego Supercomputer Center, the architecture allows us to expose additional datasets in the GLW portal if they are hosted in a manner that is accessible to the GLW's web service-based architecture. Likewise, the GLW currently only takes advantage of the ASU GEON compute nodes for DEM generation and analysis. However, the GLW architecture allows us to harness additional distributed computing resources if they become necessary.

The flexibility and extensibility of the GLW architecture lends itself to a community-oriented approach to distribution and processing of LiDAR point cloud datasets. This model allows funding agencies or data hosts to maintain control over archiving and management of their datasets, yet still allows their users to seamlessly take advantage of processing tools and resources hosted elsewhere. We propose that the GLW architecture could be adopted by the scientific community to provide a centralized data distribution and processing system that could significantly enhance user's ability to incorporate these datasets into their research (Figure 2.10).

FUTURE WORK

The GEON LiDAR Workflow as presented here represents one step in the ongoing evolution of this tool. As figure 2.6 shows, there are many pathways within the GEON Conceptual LiDAR Workflow that have not yet been implemented. Efforts are underway to enhance the GLW's functionality on a number of fronts. In addition to hosting and serving more datasets, we are actively seeking to improve DEM generation efficiency and to add new processing and analysis tools.

One area of future development is the integration of raster imagery (aerial photography or hyperspectral imagery) into the GLW workflow. It is becoming increasingly common for ultra high-resolution raster imagery to be acquired at the same time as LiDAR data and therefore it is logical to integrate these two datasets so that they could be accessed by GLW users. Because of the terabytes of data associated with raster

imagery, a geoinformatics approach to the archiving and distribution of these data is necessary.

We also recognize a need for point cloud processing tools that can be used to manipulate the point data before the user generates their DEMs. We are currently exploring a “projection engine” that would allow the user to project their point cloud data selection into a variety of geospatial coordinate systems. This functionality is essential to allow users to integrate LiDAR data with their preexisting geospatial datasets. Also necessary are a suite of tools that would allow users to interactively filter the raw point cloud to extract vegetation and bare earth products. This is important for datasets that were not classified by the provider, or that show evidence of classification errors.

Digital Elevation Model generation is a central component to the GLW and therefore we are investing considerable effort to provide additional interpolation and binning algorithms for the generation of DEMs. Because each algorithm generates a terrain model from the point data differently, it is important to offer the user a variety of choices within the GLW environment. Ultimately, we would like to offer Triangular Interpolation (TIN), Kriging, and binning methods such as local mean or local minimum in addition to the spline algorithm already implemented in the GLW (see Appendix I for more on interpolation algorithms). Our goal is also to optimize the performance of the algorithms so that the GLW can quickly generate DEMs for very large point cloud datasets.

Another logical area of development for future versions of the GLW is in tools that allow users to analyze the DEMs they produce. Currently we calculate three basic

geomorphic metrics: slope, aspect and profile curvature as derived products from the DEMs, but efforts are underway to offer additional basic functionality such as hillshade generation. Because of the computing resources available to the GLW, certain computationally difficult analysis tasks such as generating hydrologically corrected DEMs lend themselves to implementation in the GLW.

Finally, we plan to provide a greater selection of download product formats so that GLW outputs are compatible with a wide variety of software packages. Beyond the standard data formats, we are also exploring the generation of products that allow the user to visualize their data products in three-dimensions within the web browser. This functionality may include generation of vrml models or fledermaus scene files and would interface with web-based visualization efforts that GEON is pursuing for a number of research projects.

CONCLUSIONS

High-resolution topography from aerial LiDAR data is one of the most powerful new tools for the study of the earth's surface and overlying vegetation. These datasets have great utility for a wide variety of earth science, ecology and engineering applications. Access to raw LiDAR point cloud data is important for scientific research so that users can take full advantage of the information contained within these massive datasets. The size of these publicly available datasets presents a significant challenge to distribute to users. Once users do gain access to the data, they face significant software,

hardware and expertise requirements to generate digital elevation models and perform analysis of these models.

Using cyberinfrastructure available through the GEON project, we have developed an interdisciplinary, geoinformatics-based, approach to the distribution and processing of aerial LiDAR point cloud data. The proof-of-concept demonstration of this approach, the GEON LiDAR Workflow, is an end-to-end data distribution and processing workflow that leaves the computationally challenging data processing to distributed resources provided by the GEON network. As a result, the GLW has the power to democratize access to LiDAR point cloud for the greater scientific community. The GLW architecture could be adopted to build a centralized data distribution and processing resource for LiDAR data in the earth sciences.

The interdisciplinary collaboration and the utilization of cyberinfrastructure and information technology used to develop the GEON LiDAR Workflow represent an excellent model for building internet-based tools for the earth sciences. This approach to developing community data archives and processing tools that integrate distributed databases, tools and computing resources is the future of earth science research and education. Resources such as the GLW can help to facilitate new and exciting interdisciplinary science because users are more easily able to explore and integrate data. As earth science datasets become larger and more complex, a geoinformatics approach will be necessary to develop the tools and resources that facilitate data access and processing for the next generation of earth science data.

REFERENCES CITED

- American Society for Photogrammetry and Remote Sensing (ASPRS), 2005, ASPRS
 LIDAR Data Exchange Format Standard Version 1.1:
<http://www.lasformat.org/documents/ASPRS%20LAS%20Format%20Documentation%20-%20V1.1%20-%202003.07.05.pdf>
- Baltsavias, E.P., 1999, Airborne laser scanning: existing systems and firms and other
 resources: ISPRS Journal Of Photogrammetry And Remote Sensing v. 54 no. 2-3,
 p. 164-198
- Bevis, M., Hudnut, K., Sanchez, R., Toth, C., Grejner-Brzezinska, D., Kendrick, E.,
 Caccamise, D., Raleigh, D., Zhou, H., Shan, S., Shindle, W., Yong, A., Harvry, J.,
 Borsa, A., Ayoub, F., Elliot, B., Shrestha, R., Carter, B., Sartori, M., Phillips, D.,
 Coloma, F., Stark, K., 2005, The B4 Project: Scanning the San Andreas and San
 Jacinto Fault Zones: Eos Trans. AGU, 85(47), Fall Meet. Suppl., Abstract H34B-
 01.
- Carter, W.E., Shrestha, R.L., Tuell, G., Bloomquist, D., and Sartori, M., 2001, Airborne
 Laser Swath Mapping Shines New Light on Earth's Topography: Eos,
 Transactions, American Geophysical Union, v. 82. p. 549-550, 555.
- Crosby, C.J. and Arrowsmith J R., 2006, Utilization of LiDAR / ALSM Point Cloud Data
 for Earthquake Geology and Tectonic Geomorphic Mapping, Analysis, and
 Visualization: Proceedings of the 100th Anniversary of the 1906 Earthquake
 Conference
- Espinosa, A., Stennett, T., Ashenfelter, H., 2006, LiDAR Technology at Your Fingertips,

- Introducing the World's Largest LiDAR Resource Exchange Marketplace:
Imaging Notes, v. 21, no. 2, p. 16.
- Haala, N. and Breener, C., 1999, Extraction of buildings and trees in urban environments:
ISPRS Journal of Photogrammetry & Remote Sensing v. 54, p. 130-137.
- Harding, D., 2006, Overview of NASA Airborne and Satellite Laser Altimeters and
Commercial Analysis Software: Presentation given at the 2006 UNAVCO
Science Workshop Special Interest Group, "LiDAR Data: Management,
Processing, and Access".
- Haugerud, R. A., Harding, D. J., Mark, L. E., Zeigler, J., Queija, V., Johnson, S. Y.,
2004, Lidar measurement of topographic change during the 2004 eruption of
Mount St. Helens, WA: Eos Trans. AGU, 85(47), Fall Meet. Suppl., Abstract
V53D-01A.
- Haugerud, R.A., and Harding, D.J., 2001. Some algorithms for virtual deforestation
(VDF) of LIDAR topographic survey data: International Archives of
Photogrammetry and Remote Sensing, Vol. XXXIV-3/W4, pp. 211–217.
- Hudnut, K. W., Borsa, A., Glennie, C. and Minster, J.-B., 2002, High-Resolution
Topography along Surface Rupture of the 16 October 1999 Hector Mine,
California, Earthquake (Mw 7.1) from Airborne Laser Swath Mapping: Bulletin
of the Seismological Society of America; v. 92; p. 1570-1576.
- Jaeger-Frank, E., Crosby, C.J., Memon, A., Nandigam, V., Arrowsmith, J R., Conner, J.,

- Altintas, I., Baru, C., 2006, A Three Tier Architecture for LiDAR Interpolation and Analysis, *Lecture Notes in Computer Science*, v. 3993, p. 920-927, DOI: 10.1007/11758532_123.
- 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
- Lefsky, M.A., Cohen, W.B., Parker, G.G. and Harding, D.J., 2002, Lidar remote sensing for ecosystem studies: *Bioscience* v. 52, p. 19-30.
- Ludäscher B., Altintas I., Berkley C., Higgins D., Jäger-Frank E., Jones M., Lee E.A., Tao J., Zhao Y., 2005, Scientific Workflow Management and the Kepler System, In: *Concurrency and Computation: Practice & Experience*, Special Issue on Scientific Workflows.
- Maas, H.-G. and Vosselman, G., 1999, Two algorithms for extracting building models from raw laser altimetry data: *ISPRS Journal of Photogrammetry & Remote Sensing* v. 54, p. 153–163.
- Mitas, L. and Mitasova, H., 1999, Spatial Interpolation, In: P. Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind (eds), *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley, pp. 481-492.
- Mitasova, H. and Mitas, L., 1993, Interpolation by Regularized Spline with Tension I, *Theory and Implementation: Mathematical Geology*, v.25, p. 641-655.
- Mitasova, H. and Hofierka, J., 1993, Interpolation by Regularized Spline with Tension II,

Application to Terrain Modeling and Surface Geometry Analysis: *Mathematical Geology* v.25, p. 657-667.

Neteler, M. and Mitasova, H., 2004, *Open Source GIS: A GRASS GIS Approach:*

Second Edition, *Kluwer International Series in Engineering and Computer*

Science, 773, Kluwer Academic Press / Springer, Boston, Dordrecht, 424 pages.

North Carolina Floodplain Mapping Program, 2000, *About the North Carolina Floodplain*

Mapping Program: <http://www.ncfloodmaps.com/pubdocs/NCFPMPHndOut.htm>

Roering, J.J., Kirchner, J.W. and Dietrich, W.E., 1999, Evidence for nonlinear, diffusive

sediment transport on hillslopes and implications for landscape morphology:

Water Resources Research, Vol. 35, pp 853-580.

Sallenger, A. H., Krabill, W., Swift, R., Brock, J., List, J., Hansen, M., Holman, R. A.,

Manizade, S., Sontag, J., Meredith, A., Morgan, K., Yunkel, J. K., Frederick, E.,

and Stockdon, H., 2003, Evaluation of airborne scanning lidar for coastal change

applications: *J. Coastal Research*, v.19, n.1, p. 125-133.

Shrestha, R.L., Carter, W.E., Lee, M., Finer, P. and Sartori, M., 1999, *Airborne Laser*

Swath Mapping: Accuracy Assessment for Surveying and Mapping Applications:

Journal of American Congress on Surveying and Mapping, v. 59, p. 83-94.

Sinha, A.K., 2000, *Geoinformatics: A Defining Opportunity for Earth Science Research:*

White Paper Submitted to the National Science Foundation,

(<http://www.geoinformaticsnetwork.org/whitepaper.html>)

Sithole, G., and Vosselman, G., 2004. Experimental comparison of filter algorithms for

bare-Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry & Remote Sensing*, 59: 85-101.

Stoker, J.M., Greenlee, S.K., Gesch, D.B, Menig, J.C., 2006, CLICK: The new USGS Center for LiDAR Information and Knowledge: *Photogrammetric Engineering & Remote Sensing*, v. 27, p. 613-616.

U.S. Army Topographic Engineering Center Topography, Imagery and Geospatial Research Division Data Representation Branch, 2006, Survey of Terrain Visualization Software: (<http://www.tec.army.mil/TD/tvd/survey/index.html>).

Wehr, A. and Lohr, U., 1999, Airborne laser scanning---an introduction and overview, *ISPRS Journal Of Photogrammetry And Remote Sensing* v. 54 no. 2-3,p. 68-82.

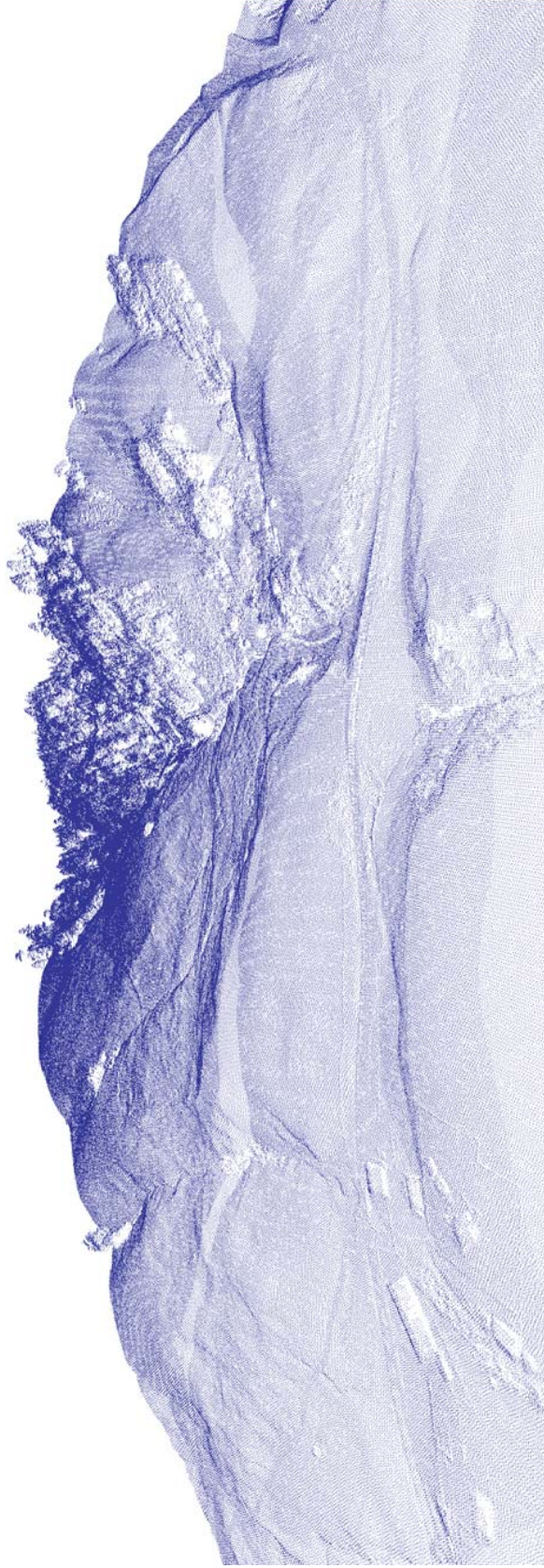


Figure 2.1. Three dimensional rendering (looking NE) of approximately 1.1 million unfiltered LiDAR point returns from the Northern San Andreas Fault dataset (Figure 2.5) near Fort Ross, CA. Note the detail with which the stands of trees in the background, landforms and buildings (bottom left) are represented by the point cloud data. U.S. Highway one passes through the middle of the image.

Figure 2.3. Generalized aerial LiDAR acquisition and processing workflow consisting of four steps: 1) Aerial data acquisition—scanner, IMU, and GPS (modified from R. Haugerud, USGS: http://duff.geology.washington.edu/data/raster/lidar/About_LIDAR.html)

2) processing of laser ranging, GPS and IMU data to generate LiDAR point cloud (modified from Harding, 2006), 3) generation of classified point cloud, and 4) generation of digital ground and vegetation models. Point cloud classification and generation of digital ground and vegetation models directly controls the specifications of the products upon which analyses will be performed.

Figure 2.4. Vertical and three dimensional view of full feature (unfiltered) LiDAR digital elevation model (top). Brown shades show ground surface. Vegetation is shown in shades of green with vegetation color coded by canopy height (darker is taller). Lower images show the digital elevation model produced from just the point returns classified as having been returned from the ground surface. Images are from a portion of the Northern San Andreas Fault (NSAF) LiDAR dataset at Anapolis Road near Sea Ranch, CA (middle area of figure 2.5). Note the linear trace of the NSAF clearly visible in the bare earth digital elevation model. The fault's obvious geomorphic expression is obscured by the vegetation in the upper images. The ability to 'virtual deforest' the landscape is extremely powerful for many geoscience applications (e.g. Haugerud and Harding, 2001).

Figure 2.5. Map showing the extent of the Northern San Andreas Fault (NSAF) LiDAR dataset in Sonoma and Mendocino Counties, California (shown in orange). This dataset is one of three pilot datasets for the development of the GEON LiDAR Workflow. The NSAF data were acquired in February 2003 by Terrapoint LLC with funding from NASA in collaboration with the U.S. Geological Survey. LiDAR data were acquired in an approximately 418 square mile area with a focus on the NSAF and associated marine terraces. With approximately 1.2 billion LiDAR classified returns, this dataset is a good example of the data types and volumes that are common in aerial LiDAR datasets.

Figure 2.7. A) Screen capture of the GEON LiDAR Workflow portal. Red box shows the extent of (B) and orange box shows the extent of (C). B) Detail view of the spatial and attribute selection portion of the portal. Users can use the WMS (Web Mapping Service) map to zoom to the portion of data of interest and then draw a box on the map to select a subset of data. If attribute information for the dataset is available, that selection can be made here as well. C) Detail view of the processing portion of the portal. Users choose their products (DEM and derived products) as well as the file format for those products. Processing parameters such as DEM resolution and spline interpolation settings (Mitasova and Mitas, 1993) can also be defined.

Figure 2.10. Proposed GEON-Based Model for accessing and processing of distributed community LiDAR datasets. Although the organizations named in the figure are not committed, they are exemplary of different agencies and research organizations likely to fund publicly available LiDAR datasets. In this model, a single web-portal allows users to seamlessly access datasets hosted and maintained by various organizations and agencies. Likewise, various tiers of distributed computing resources can be recruited depending upon the size of the processing job submitted, ranging from a single computing cluster to the TeraGrid (“the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research”. <http://www.teragrid.org/about/>). The orchestration of the various databases, processing tools and computing resources would be managed by the cyberinfrastructure developed by GEON and currently implemented in the GEON LiDAR Workflow. Red arrows show the pathways through this model that are implemented as the GLW, future pathways are shown in black.

APPENDIX I

EXPLORATION OF LIDAR POINT CLOUD DATA ARTIFACTS, RETURN DENSITY AND DIGITAL ELEVATION MODEL GENERATION

SUMMARY

The growing availability of LiDAR (Light Distance And Ranging (a.k.a. ALSM – Airborne Laser Swath Mapping)) data in the earthquake geology and tectonic geomorphology communities (among numerous others – see Introduction and Chapter 2) means that these powerful data are being utilized in an increasing number of research projects. LiDAR point cloud data (x, y, z, return classification) (Figure I.1) are challenging to manipulate, so users typically only take advantage of interpolated surfaces (digital elevation models; DEMs) generated by the LiDAR data vendor for their analysis. However, by not returning to the LiDAR point cloud data, users may fail to fully exploit the richness of these data sets.

Initiating geomorphic analyses and visualizations with the point cloud gives users more understanding of the data and control over how those data characterize the landscape. Details such as the interpolation algorithm and grid resolution can significantly affect the manner in which the resulting DEM represents the landscape. In addition, beginning with the LiDAR point cloud data allows the user to assess the homogeneity and density of ground returns in the area of interest (Figure I.2) and to evaluate potential artifacts in the data caused by errors in data acquisition and processing (e.g. point misclassification) (Figure I.1). By understanding the variation in ground-return density (which can vary due to topography, canopy characteristics and acquisition

parameters) (Figure I.5), the user has a better understanding of potential artifacts that may be introduced into their DEMs by this variation.

Using LiDAR point cloud data from the Northern San Andreas Fault (Figure 2.5) recently made available via the GEON LiDAR Workflow (GLW) (<http://www.geongrid.org/science/lidar.html>) (also see Chapter 2 in this volume), I focus on evaluation of LiDAR point return density in both forested and unforested landscapes near Fort Ross, California. This prototypical methodology was originally presented as Crosby et al., 2006 and demonstrates that by evaluating LiDAR ground return density before generating a DEM, the user is able to performed “informed” DEM generation whereby the resolution of the DEM is optimized to take full advantage of the data resolution (Figure I.3). This analysis revealed that DEMs generated from LiDAR point cloud data often do not take full advantage of the point return density, producing DEMs that have multiple returns per DEM cell. I propose that LiDAR DEMs can be accurately produced down to a one point per pixel on average threshold. In this appendix I also explore the accuracy of common DEM interpolation algorithms to fit the LiDAR point cloud (Figure I.4). In areas of high-ground return density relative to the grid resolution, there is little variation between interpolation algorithms in the accuracy with which they fit LiDAR ground returns.

Through interactive exploration and interpolation of LiDAR point cloud data, users gain a better understanding of the strengths and weakness of their data and are able to optimize DEM generation to represent the landscape they are studying. This approach

to working with LiDAR data allows tectonic landforms to be delineated more efficiently and with greater detail than by working with the vendor generated DEMs.

REFERENCES CITED:

Crosby, C.J. and Arrowsmith J R., 2006, Utilization of LiDAR / ALSM Point Cloud Data for Earthquake Geology and Tectonic Geomorphic Mapping, Analysis, and Visualization: Proceedings of the 100th Anniversary of the 1906 Earthquake Conference

Figure I.1. 3D visualization of classified LiDAR point cloud data. These images show a NE view up Mill Gulch near Fort Ross, CA. Note U.S. Highway One in the lower left of each image. A) 3D rendering of ground returns (black) and vegetation returns (green) over the bare earth DEM (brown). B) 3D rendering of ground returns (black) and vegetation DEM (shades of green color coded by height with darker shades of green for taller vegetation -- maximum tree height is approximate 280 ft) over the bare earth DEM (brown). C) 3D rendering of ground returns (black) over the bare earth DEM (brown). Note the lack of ground returns (gaps in the black points) underneath the canopy. D) 3D rendering of vegetation returns (green) over the bare earth DEM (brown). Note returns classified as vegetation on the flat surfaces in the foreground. These points are misclassified, reclassifying them as ground returns would yield higher ground return densities in these areas. The misclassification of these points is likely an error due to swath mismatches within the processed data set. This offset is likely traceable to the mislocation of the aircraft due to GPS or Inertial Navigation problems.

Figure I.1

Figure I.2. LiDAR return density evaluation. This figure shows the number of LiDAR ground (A) and vegetation (B) returns for each 6 ft pixel in the DEM. Note the limited vegetation returns on the flat, grass-covered, marine terraces and high return density (up to 12 per pixel) in the heavily forested drainages. Also note the very low number of ground returns (< 1 per pixel) beneath the forest canopy (recall Figure I.1C). This lack of ground returns beneath dense vegetation is likely due to poor penetration of the LiDAR pulse through the canopy and the point classification algorithm. The plots show the distribution of points per pixel for both ground (C) and vegetation (D) returns. Note that the majority of 6 ft pixels have 2 or more ground returns. Thus, in certain portions of the data set, it may be appropriate to generate DEMs at resolutions better than 6 ft in order to take full advantage of the richness of these data. Conversely, these figures illustrates that in certain areas, due to poor penetration of the tree canopy, there are very few ground returns. In these areas the density of ground returns may not support the 6 ft DEM resolution as produced by the data vendor. If DEM generation is going to be undertaken in these areas of very low return density, it is necessary to interpolate to span the gaps in the data and thus the method (gridding algorithm) becomes important.

Figure I.2a&b

Figure I.2c&d

Figure I.3. Testing digital elevation model resolution based on LiDAR ground return density. In order to illustrate how an understanding of ground return density can inform the generation of digital elevation models (DEMs) from LiDAR point cloud data, a ~280,000 ground return subset of the Northern San Andreas LiDAR data was selected.

A) The dataset extent is shown in the yellow box. This region was selected because it is largely unvegetated and has a high density of ground returns (B). In addition, the area is crossed by a marine terrace riser and also contains a number of sea stack remnants. With the exception of the brush in the upper portion of the sample area, the point density evaluation (B) shows that each 6 foot DEM pixel is sampled by at least 2 LiDAR ground returns. This density of ground returns suggests that a DEM at resolutions greater than 6 feet are supported by the data. C) Standard, vendor provided 6 foot DEM for the area shown in (B). D) Point cloud data interpolated to produce a 3 foot DEM. Note the increased clarity of subtle features in the landscape when compared to the 6 ft DEM (C). E) Point cloud data interpolated to produce a 1.5 foot DEM. Again, note the increased clarity of subtle landscape features when compared to both the 6 (C) and 3 foot (D) DEMs. Given the ground return density evaluation shown in (B), the 1.5 foot DEM is likely at the threshold of what is appropriate for this data set.

Figure I.3a&b

Figure I.3c,d&e

Figure I.4. Comparison of common DEM generation interpolation algorithms. The images in this figure explore the accuracy of common DEM interpolation algorithm to fit the LiDAR point cloud. The analysis is performed for the region shown in the yellow box in Figure 3A. Each image is a 3 foot DEM produced with one of four interpolation algorithms: A) Spline B) TIN (Triangular Interpolation Network) C) IDW (Inverse Distance Weighted) D) Kriging. The elevation of the grid at each of the 280,000 individual ground return points was then extracted from the DEMs to test the fit of the surface to the original ground returns. The histograms show the differences between the points and the corresponding DEM pixel. The mean and standard deviation are also indicated. This analysis indicates that in areas of high-ground return density (like this sample data set), there is little variation between interpolation algorithms in the accuracy with which they fit LiDAR ground returns. Therefore, interpolation algorithm is likely not the most important factor in the generation of DEM in such situations.

Figure I.4a&b

Figure I.4c&d

Figure I.5. *Digital elevation model generation in areas of low ground return density.*

A) Hillshade of the first return surface (full feature) DEM for a portion of the Northern San Andreas fault (NSAF) LiDAR data (Figure 2.5) that is characterized by very dense forest canopy on steep slopes. B) Hillshade of the bare earth DEM of the area shown in (A). The San Andreas fault zone crosses the image from the upper left to lower right (with in the orange box). Much of the NSAF is located in landscapes like this example. Although the LiDAR data helps to reveal the geomorphology of the SAF fault zone, the density of ground returns in these areas is low due to difficulty of LiDAR penetration through the dense forest canopy. When these data are interpolated, the sparse density of ground returns results in artifacts such as the triangular facets (associated with the TIN interpolation used by the data provider) visible in the bare earth image. Changing interpolation algorithm and/or parameters may enhance the landscape representation in these areas by altering the way that the interpolated surface treats the sparse ground returns. C) This image shows the number of ground returns per 6 foot DEM pixel. Note that in most of the forested areas there is less than one ground return per pixel. Interpolating these data to DEMs of resolutions less than 6 feet would not be appropriate in this case. D) Hillshade of a 6 foot bare earth DEM of the area shown in (A). This DEM was generated by interpolating the same ground return LiDAR point cloud data (~301,000 points) as in (B) but the spline algorithm (settings: DMin=1 ft, tension=40, smoothing=0.1) available in the GEON LiDAR workflow was used instead of a TIN. The spline algorithm does not produce the large facets associated with the TIN so it provides a smoother representation of the landscape. The spline surface is characterized

by generally smooth surface with occasional bumps or “pimples”. However, the extent to which the representation of the SAF fault zone geomorphology is enhanced by the spline is open to debate. Furthermore, the triangular facets artifact produced by the TIN are a clear indication of low ground return density. These facets alert the users to the return density problem with the data set while the spline algorithm conceals some of these data density issues and may therefore fool the user into believing that the data set is robust. One could also use a semi-transparent overlay color coded by return density to assess this issue. For the reasons discussed above, users working with these data may wish to further explore the role of interpolation algorithm and parameters in low-ground return environments.

Figure I.5a&b

Figure I.5c&d

APPENDIX II

EXPLOITING LIDAR DATA FOR REGIONAL MORPHOLOGIC CORRELATION AND DATING OF WAVE-CUT AND FAULT-CONTROLLED LANDFORMS

SUMMARY

The capability to generate high-resolution Digital Elevation Models (DEMs) from LiDAR data (Light Distance and Ranging, also known as Airborne Laser Swath Mapping, or ALSM) across broad geographic regions provides a new tool for studying landscape response to tectonic deformation. Expanded LiDAR coverage from GeoEarthscope (http://facility.unavco.org/project_support/es/geearthscope/) and the National Center for Airborne Laser Mapping (NCALM) (<http://www.ncalm.ufl.edu/>) offers the prospect of applying these data to a variety of tectonic geomorphic studies. The data volume and point-density of LiDAR allows extensive repetition of profile-based landscape analyses without the need for laborious survey transects. Traditional DEMs, such as those available through the USGS National Elevation Dataset, lack the resolution necessary for these types of analyses. In this appendix I develop a scheme for exploiting LiDAR data for landform correlation by conducting profile-based morphologic dating (using linear diffusion) of fault scarps and marine, lake and fluvial shorelines. The resolution and geographic extent of LiDAR coverage makes broad spatial correlations possible, assuming that controls on the hillslope processes are relatively constant across the region. Due to the high data density and their digital format, numerous topographic profiles can be extracted from a DEM and analyzed for morphologic age. Correlation from profile to profile can then be established by comparing morphologic age for various

landforms in a research area. With calibration, morphologic dating also offers the opportunity to constrain absolute ages of landforms. Once calibrated, landforms across the region can be quickly dated via profile-based analysis of the LiDAR-derived DEM. In addition, morphologic comparison of landforms of known age offers the opportunity to test the role of other constraints, such as aspect, microclimate, and substrate type on landform development by diffusive processes.

The linear hillslope diffusion method requires the simplistic assumption that controls on hillslope processes are relatively constant cross the region of correlation. Other assumptions include: transport-limited conditions, regolith transport rate increases with increasing slope, a simple ramp-shaped initial topography, and no geomorphic transport in or out of the strike of the profile. Simple linear diffusion is expressed as the following analytical solution for an initial ramp-shaped form (Andrews and Hanks, 1985) (Figure 1):

Equation 1.

$$H(x, t) = (\theta - b) \left(\frac{\kappa t}{\pi} \right)^{1/2} \left\{ \exp \left(-\frac{x + a/(\theta - b)}{4\kappa t} \right)^2 - \exp \left(-\frac{x - a/(\theta - b)}{4\kappa t} \right)^2 \right\} \\ + \frac{\theta - b}{2} \left\{ \left(x + \frac{a}{\theta - b} \right) \operatorname{erf} \left(\frac{x + a/(\theta - b)}{(4\kappa t)^{1/2}} \right) - \left(x - \frac{a}{\theta - b} \right) \operatorname{erf} \left(\frac{x - a/(\theta - b)}{(4\kappa t)^{1/2}} \right) \right\} \\ + bx$$

Where:

b = far field fan slope

a = scarp half offset

k = diffusion constant ($\text{m}^2/1000 \text{ yrs}$)

t = time

θ = initial scarp slope

H = elevation along the profile which is a function of distance (x) and time (t)

Analysis of scarp profiles was performed using the Diffusion Scarp Dater MATLAB graphical user interface written by George Hilley (Hilley and Arrowsmith, 2001). This tool allows calculation of finite scarp RMS, by comparing the observed profiles from LiDAR to the modeled ones, and forward modeling of topographic transect data. The tool is available for download at: <http://activetectonics.asu.edu/diffuse>.

In this appendix, I first demonstrate a methodology for performing morphologic age dating on synthetic profiles. It illustrates the ability of morphologic dating to differentiate landforms of morphologic age 50 m^2 from ones of 100 m^2 . Next, I revisit a pair of classic morphologic dating studies (Hanks et al., 1984; Hanks and Wallace, 1984) to test the results of my methodology against the published results for these studies. Finally, I apply our methodology to LiDAR data sets from marine terraces along the Mendocino County, California coast and from normal fault scarps in Death Valley, California.

Application of our morphologic dating technique to these recently acquired California LiDAR data sets illustrates the potential power of morphologic dating for establishing correlations among regional landforms. However, this analysis also reveals

that this kind of morphologic dating and analysis is over simplified and is highly dependent on transect selection. Typically, transect selection has been done by the geomorphologist's qualitative assessment in the field. The abundance of high-resolution topographic data provided by LiDAR creates a new suite of geomorphic and technical complications that need to be addressed in order to make morphologic correlation effective. Ultimately, these complications will provide insight into the geomorphic process once fully understood. The complications in the morphologic age analysis can be traced to following causes: landform rejuvenation, non-transport limited conditions, non-linear diffusion, and non-diffusive conditions. Also, high-frequency noise in the LiDAR data effects slopes along the profile. Undersampling by using every other DEM cell along the profile for the slope calculations help to reduce this problem.

Figure II.1. *Simple scarp diffusion: finite slope initial form.* Model of simple linear diffusion of a theoretical scarp-like landform with diffusion constant (k) of $10 \text{ m}^2/\text{ka}$. Our algorithm assumes initial vertical riser morphology rapidly evolves to a steep, ramp-shaped topography via mass wasting processes immediately after formation (θ in equation 1). Diffusive process then continue to modify the riser (e.g. Rosenbloom and Anderson, 1994; Hanks et al., 1984, Hanks, 2000).

Figure II.1.

Figure II.2. *Illustration of morphologic dating method on synthetic "LiDAR data".*

A) Map view of theoretical field area. In this scenario, two terrace risers are offset right-laterally by a dextral strike-slip fault. Five profiles from each of the terrace risers were analyzed to test their correlation across the fault. B) Synthetic profiles produced by forward model calculations of 500 and 1000 m² profiles starting with a 10 m riser and flat tread (for a diffusion constant (k) of 10 m²/ka the morphologic age (*kt*) yields an absolute age of 50,000 yrs and 100,000 yrs respectively). To the resulting profiles we added +/- 50 cm of noise to simulate local heterogeneity in the surface as is typically encountered and would be likely in the LiDAR derived profiles. C) Best fitting model profiles to the synthetic "data" at left. Initial scarp morphology shown as blue line, topographic "data" are red dots. D) Illustration of the relationship between RMS error and morphologic age for the five different profiles of the two different risers. This analysis demonstrates the ability of morphologic dating to differentiate landforms of morphologic age 500 m² from ones of 1000 m².

Figure II.2a&b

Figure II.2c&d

Figure II.3. Hanks et al., 1984 - Profile modeling of the Santa Cruz, CA marine terrace risers revisited. Using the topographic profiles published in Hanks et al., 1984, I recreated this Santa Cruz terrace study to test my methodology and demonstrate the technique's utility for morphologic dating of marine terrace risers. A) Santa Cruz, CA marine terrace topographic profile recreated from Hanks et al., 1984. Ages for the three risers come from U-Th and amino acid racemization data, global sea level curves and the assumption of a constant uplift rate of 0.35 m/ka (see Hanks et al., 1984). B) Plot showing slope calculated along the marine terrace topographic profile. Profiles illustrate qualitative observation that the older terrace risers have more subtle topography. The peak slope diminishes and the scarp widens. C) Diffusion model calculations (solid lines) for the Santa Cruz terrace risers. Assumed initial morphology shown as dashed blue line. Actual topography is shown as red dots. Model parameters are summarized in (D). D) Table showing model calculation parameters and results for both the original Hanks et al., 1984 paper and this study. My methodology yields similar morphologic ages (*kt*) to those calculated by Hanks et al., 1984 for the Santa Cruz marine terrace flight.

Figure II.3a&b

Figure II.3c

Figure II.3d

Figure II.4. *Hanks and Wallace, 1984 - Morphological analysis of Lake Lahontan shoreline scarps revisited.* From topographic profiles of Lake Lahontan high stand shorelines published in Hanks and Wallace, 1985, I revisited their quantitative comparison of profiles to demonstrate how morphologic dating can be used to correlate landforms. A) Centered Lake Lahontan shoreline topographic profiles recreated from Hanks & Wallace, 1985 (topographic profiles as published were reused). Shown are the profiles for which Hanks & Wallace performed model calculations. B) Example diffusion model calculations (solid lines) for two Lake Lahontan shorelines. The assumed initial riser morphology is shown as the dashed blue line. Actual topography is shown as red dots. Model parameters are summarized in (D). D) Table showing model calculation parameters and results for both the original Hanks and Wallace, 1984 paper and this study. My methodology yields similar morphologic ages (kt) to those calculated by Hanks and Wallace, 1984 for the Lake Lahontan terrace flight.

Figure II.4a

Figure II.4b

Figure II.4c

Figure II.5. Application of morphologic dating to LiDAR/ALSM data - Sheep Creek fan fault scarps, Death Valley, CA. In this analysis, I blindly apply the morphologic dating technique to fault scarps cutting the Sheep Creek alluvial fan (data courtesy of Dr. Thad Wasklewicz, University of Memphis). Topographic profiles were extracted from the ALSM data and then analyzed for morphologic age. A) Hillshade of 1 meter digital elevation model derived from ALSM data. Three fault scarps clearly offset the Sheep Creek fan. The locations of topographic profiles extracted from the DEM are shown as colored lines (red: scarp 1, green: scarp 2, orange: scarp 3). B) Slope-offset plot (Hanks, 2000) for topographic transects across the three Sheep Creek fan scarps. Transect data shown as colored Xs. Also shown are model calculations for a variety of kt values. The transects generally plot at low kt values ($< 5 \text{ m}^2/\text{ka}$) however the significant scatter in the data makes it difficult to associate any scarp with a single kt value. Plotting the topographic transects in the slope-offset space reveals significant variation in kt along strike for all three scarps. In general, the slope offset plot suggests low morphologic ages ($< 5 \text{ m}^2/\text{ka}$) for the Sheep Creek scarps. Although it is difficult to distinguish these profiles from one another due to the scatter, the slope-offset analysis is likely to allow these scarps to be distinguished from ones of $50 \text{ m}^2/\text{ka}$. C) Calculated model fit to a select topographic profile. Note relatively poor fit to the data in the upper and lower parts of the scarp. Many of the topographic profiles reveal an over-steepening of the scarp near its base and a bevel in the upper scarp. Qualitative forward modeling (manual selection of the best-fit) of these transects may yield a very different kt than that of the analytical solution, depending upon what portion of the scarp you choose to fit. The

over-steepened base and the upper bevel of the scarp may suggest that the scarps are a product of multiple earthquakes. Mating the morphologic dating analysis with paleoseismic trench studies may yield additional insight into the number of earthquakes responsible for the scarp morphology. D) Relationship between RMS and kt for model calculation shown in (C). Note the range of potential kt values that fit the topographic data. This range of kt values is due to the poor fit of the model calculation to the topographic profile data. E) Table showing model calculation parameters and results for the Sheep Creek alluvial fan scarps. Kt values obtained using both model calculations and qualitative best-fit for each topographic profile are shown. Note that the kt value for each scarp profile can vary dramatically depending on which method is used to fit the data.

Figure II.5a

Figure II.5b

Figure II.5c&d

Figure II.5e

Figure II.6. *Application of morphologic dating to LiDAR/ALSM data - Marine terraces, Mendocino County, CA.* Morphologic dating has potential utility for correlating marine terraces over significant lengths of coast line. Here I compare topographic transects extracted from a small piece of coastal LiDAR data to test the technique's ability to differentiate risers of different ages (location is northern most portion of Figure 2.5). A) Hillshade of 1.8 m bare earth DEM derived from LiDAR data. The three lowest marine terrace risers are mapped by colored lines (blue: riser 1, green: riser 2, orange: riser 3). Mapping is from aerial photography and the LiDAR data (Prentice et al., 2003). Topographic profiles extracted for this study are shown as straight line segments, colored by riser. B) Upper plots show centered topographic profiles across riser 1 (left) and riser 3 (right). The lower left plot shows slope along the riser 1 topographic profiles. Red lines denote one sigma buffer around the red profile to demonstrate the extent of the high-frequency noise in the gridded LiDAR DEM. Note the significant along strike variation in riser morphology. The lower right plot shows slope along the riser 3 topographic profiles. Note the along strike variation in riser morphology - risers shown with blue, yellow and cyan dots and blue x's all show evidence for rejuvenation. C) Calculated model fit to transect across riser 1 (black line). The green line shows a qualitative best fit (determined by manually selecting the best fit) to the upper portion of the riser. The analytical solution is driven to higher *kts* due to “excess mass” on the terrace platform. Excess mass is typically material that has moved into the plane of the profile by non-diffusive processes such as wind transport (see Hanks et al., 1984 for a full discussion of the excess mass problem). In order to avoid this

problem, I favor a best fitting model of the upper riser only when the mass excess problem exists on the terrace tread below. D) Table showing model calculation parameters and results for the Mendocino County, CA marine terrace risers. Kt values obtained using both model calculations and qualitative best-fit for each topographic profile are shown. Note that the kt value for each scarp profile can vary dramatically depending on which method is used to fit the data. For this example I also calculate the diffusion constant (k) for each of the transects using the assumed ages of oxygen isotope stage 5a (~83 ka) for riser 1 and oxygen isotope stage 7 (~194 ka) for riser 3 (C. Prentice, personal communication). Note that many of the transects from riser 3 yield very low morphologic ages, often younger than the riser 1, suggesting that rejuvenation of this older landform by non-diffusive processes such as fluvial erosion and anthropogenic activity is a concern.

Figure II.6a

Figure II.6b&c

Figure II.6d

REFERENCES CITED

- Andrews, D.J. and Hanks, T.C., 1985, Scarp degraded by linear diffusion: Inverse solution for age: *Journal of Geophysical Research*, v. 90, p. 10193-10208.
- Arrowsmith, J.R., Rhodes, D.D., and Pollard, D.D., 1996, Hillslope development in areas of active tectonics: *Journal of Geophysical Research*, v. 101, p. 6255-6275.
Correction: 1999, *Journal of Geophysical Research*, v. 104, p. 805.
- Avouac, J-P., 1993, Analysis of Scarp Profiles: Evaluation of Errors in Morphologic Dating: *Journal of Geophysical Research*, v. 98, p. 6745-6754.
- Hanks, T.C., 2000, The Age of Scarplike Landforms From Diffusion-Equation Analysis, In: Noeller, J.S., Sowers, J.M., Lettis, W.R. eds., *Quaternary Geochronology: Methods and Applications*: American Geophysical Union, Washington, D.C., 582 p.
- Hanks, T.C., Wallace, R.E., 1985, Morphological Analysis of the Lake Lahontan Shoreline and Beachfront Fault Scarps, Pershing County, Nevada: *Bulletin of the Seismological Society of America*, v. 75, p. 835-846.
- Hanks, T.C., Bucknam, R.C., Lajoie, K.R., Wallace, R.E., 1984, Modification of Wave Cut and Fault-Controlled Landforms: *Journal of Geophysical Research*, v. 89, p. 5771-5790.
- Hilley, G. E., 2001, Landscape development of tectonically active areas: Tempe, Arizona, Arizona State University, Ph.D. dissertation.
- Hilley, G.E., and Arrowsmith, J.R., 2001, Transport and production-limited fault scarp simulation software, in: Merritts, D., Burgman, R., *Tectonics and Topography*:

Crustal Deformation, Surface Processes, and Landforms: Geological Society of America Short Course Manual, Annual Meeting, 66 p., with CD image gallery.

Prentice, C.S., Crosby, C.J., Harding, D.J., Haugerud, R.A., Merritts, D.J., Gardner, T., Koehler, R.D., and Baldwin, J.N., 2003, Northern California LIDAR Data: A Tool for Mapping the San Andreas Fault and Pleistocene Marine Terraces in Heavily Vegetated Terrain, *Eos Trans. AGU*, 84 (46), Fall Meet. Suppl., Abstract G12A-06, 2003