A NATIONAL FRAMEWORK FOR COMMUNITY LIDAR DATASETS

CROSBY, Christopher J. , ARROWSMITH, J Ramón - School of Earth and Space Exploration, Arizona State Univ, Tempe, AZ 85281-1404
BARU, Chaitan - San Diego Supercomputer Center, Univ of California, San Diego, La Jolla, CA 92093-0505
MEERTENS, Charles - UNAVCO Facility, UNAVCO, Inc, 6350 Nautilus Drive, Boulder, CO 80301

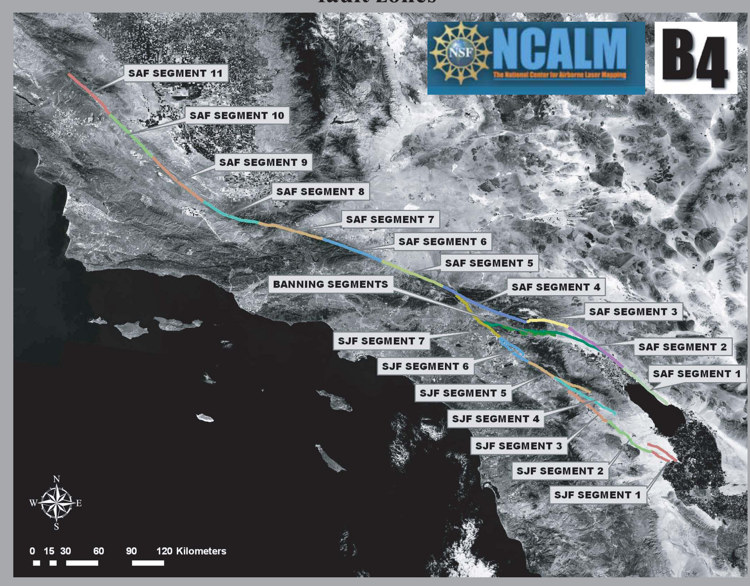
ARIZONA STATE UNIVERSITY
http://activetectonics.asu.edu
http://lidar.asu.edu

GEON CYBERINFRASTRUCTURE FOR THE GEOSCIENCES

I. INTRODUCTION & MOTIVATION

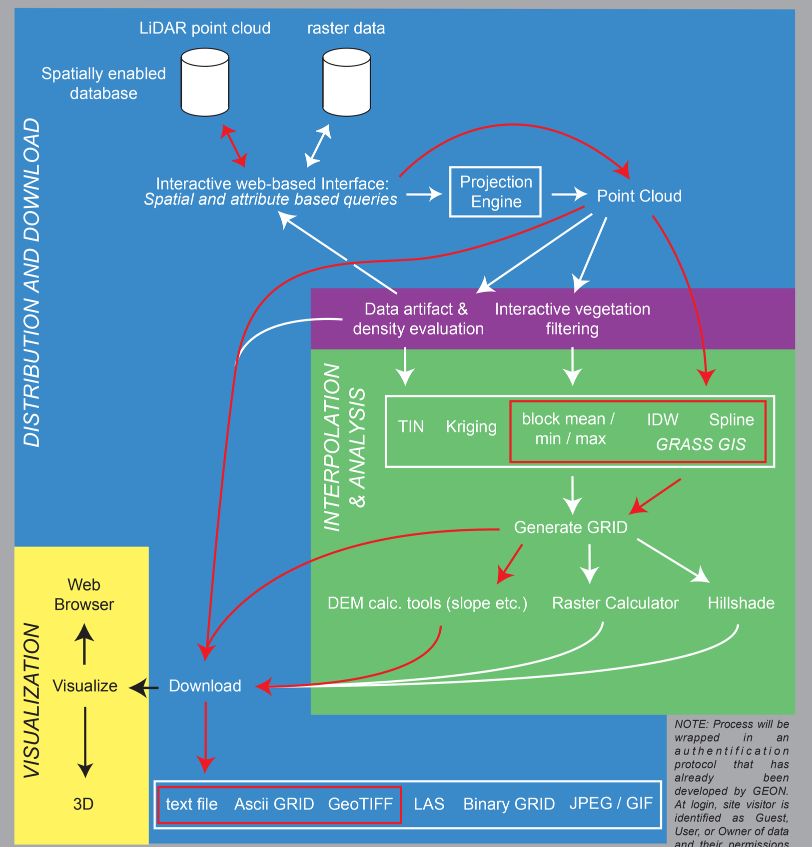
- LiDAR / ALSM generates massive data volumes - billions of returns are not uncommon in these data sets.
- Processing and analysis of these data requires significant computing resources not available to most geoscientists.
- Geoscience users typically work with digital elevation models (DEMs) generated from the LiDAR point cloud data. However, DEM generation from these data challenges typical GIS / interpolation software.
  - our tests indicate that ArcGIS, Matlab and similar software packages struggle to interpolate even a small portion of these data.
- LiDAR data are often acquired as a community resource (e.g. GeoEarthscope). Because of the large size of these datasets, a novel approach is necessary to facilitate community access to both the data as well as processing resources.

The B4 Project (an exemplary community dataset): LiDAR coverage of the Southern San Andreas and San Jacinto fault zones



- Acquired "to obtain pre-earthquake imagery necessary to determine near-field ground deformation after a future large event (hence the name B4), and to support tectonic and paleoseismic research" (Bevis et al. 2005).
- Very high-res. dataset supporting DEM generation at 25-50 cm. As a result of return density and geographic extent the dataset is massive, containing ~15 billion returns.
- Large demand from user community for access to B4 DEMs at various resolutions.

II. CONCEPTUAL GEON LiDAR WORKFLOW



- We propose utilizing the cyberinfrastructure developed by GEON to offer online data distribution, DEM generation, and terrain analysis for large LiDAR point cloud datasets.
- By using distributed computing resources, user is able to quickly run multiple jobs and compare results. We leave the computationally intensive data processing to resources available through GEON and offer user downloadable products in common file formats.
- We have developed a completely internet-based workflow that runs on the GEON grid and utilizes modular web services to complete a variety of processing and analysis tasks.
- *Goal*: Create an interactive processing environment for iteration and exploration of various interpolation and processing options. Optimize landscape representation based upon application of the data.

III. GEON WORKFLOW IMPLEMENTATION

OVERVIEW:

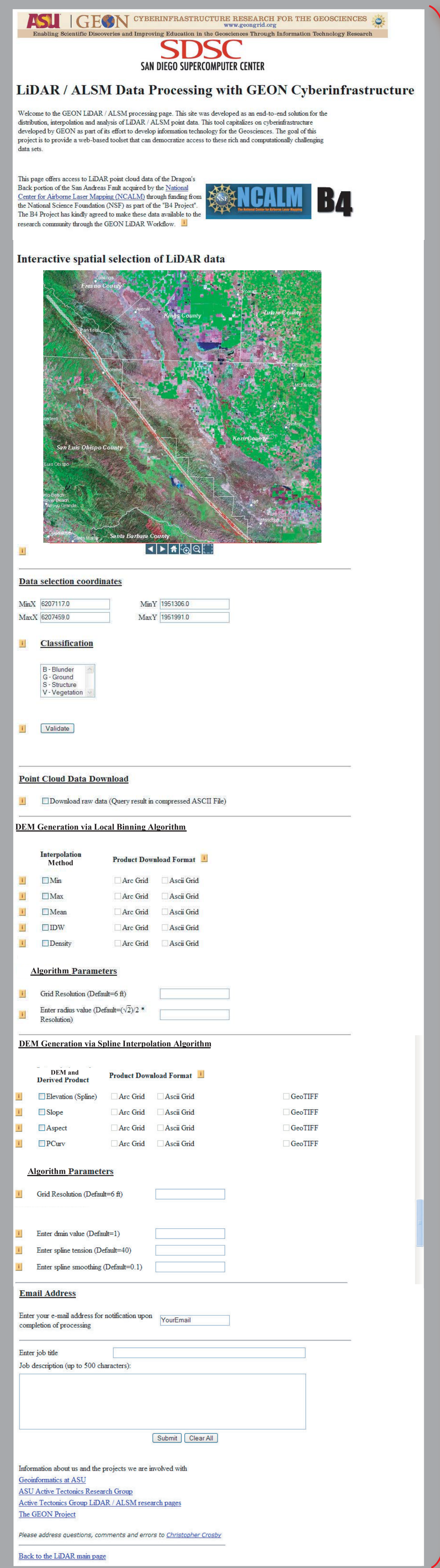
- Using GEON cyberinfrastructure we have developed an internet-based LiDAR distribution and processing (DEM and derived product generation).

Features:

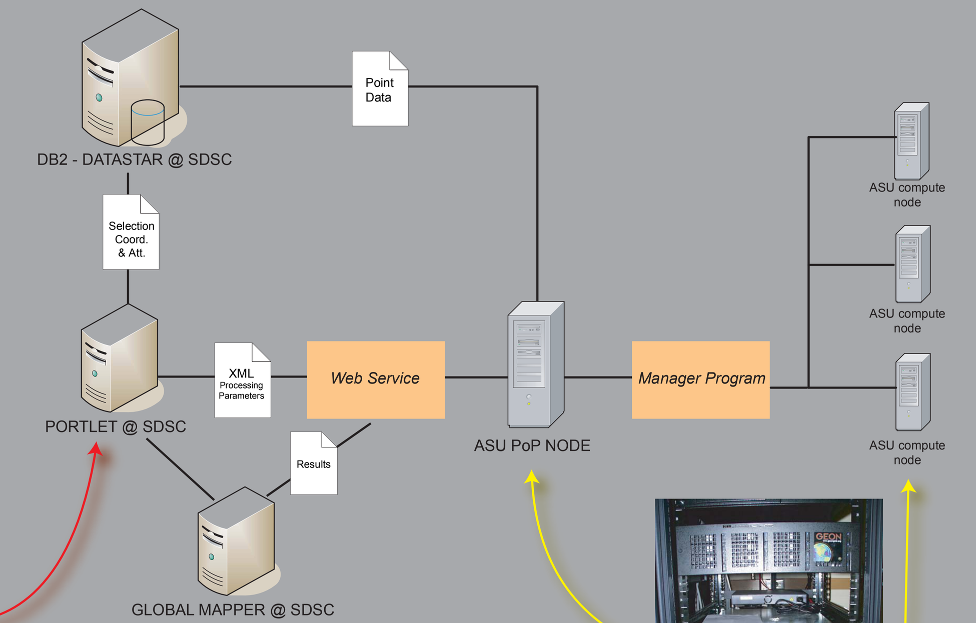
- Spatial and attribute based queries on raw LiDAR point cloud data.
- Local binning or Spline interpolation to Digital Elevation Model (DEM).
- User control over interpolation parameters.
- Slope, aspect and profile curvature (pcurv) derived products.
- Download of of products in TIFF (with world file), ASCII and ESRI Arc ASCII (Arc GRID) formats.
- Visualization of data products via web browser window or in 3D via Fledermaus free viewer iView3D (under development) and our own OpenGL-based tool "LVIZ".
- Job monitoring and archiving - including user-defined job name and description

IMPLEMENTATION:

- The GEON LiDAR workflow utilizes the Kepler System (kepler-project.org) to integrate heterogeneous local and remote tools in a single interface:
  - Web and Grid services
  - GIS services (GRASS Open Source GIS) and Local Binning Algorithm.
  - Remote tools via SSH, SCP and GridFTP
  - Relational and spatial databases (Datastar Terascale Machine at San Diego Supercomputer Center)

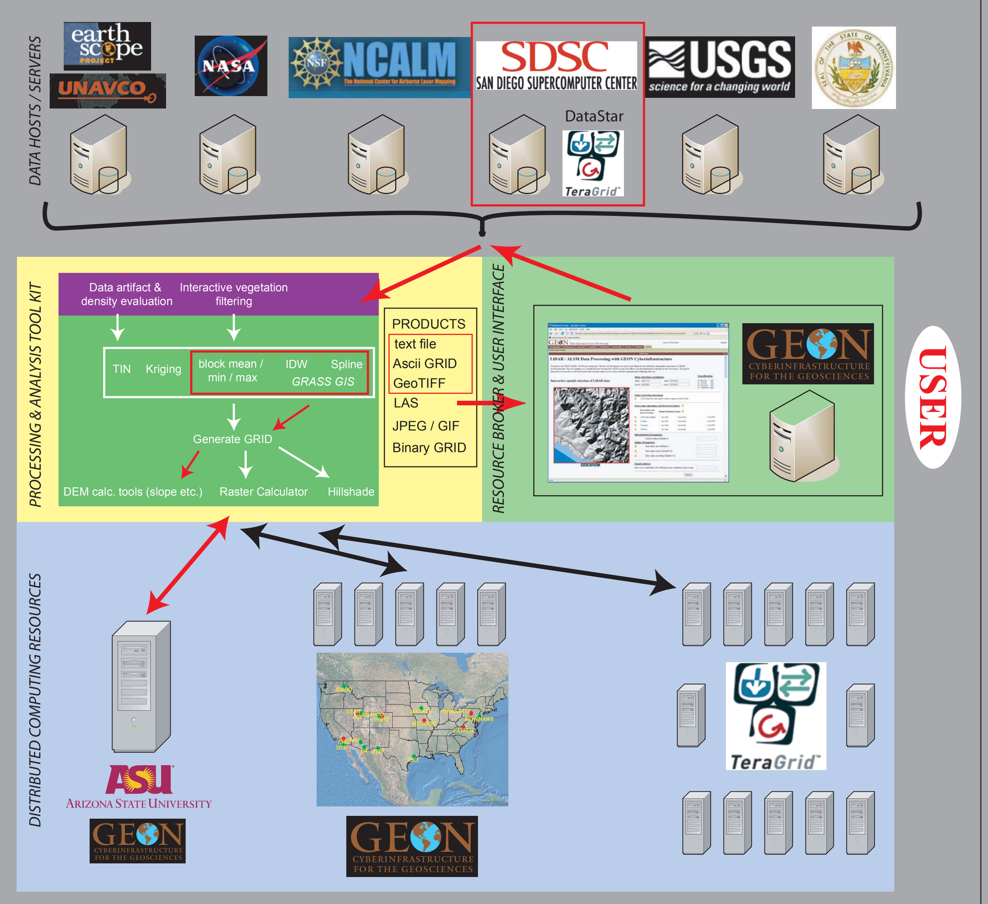


DISTRIBUTED RESOURCE CONFIGURATION:



ASU GEON node "Agassiz"- six 2.8 Ghz Intel Xeon processors w/ 2 GB RAM per CPU. Running Linux ("Rocks" based on Red Hat 7.3). 2.5 TB of storage. All GEON participating institutions have similar resources that are made available via the GEON Grid

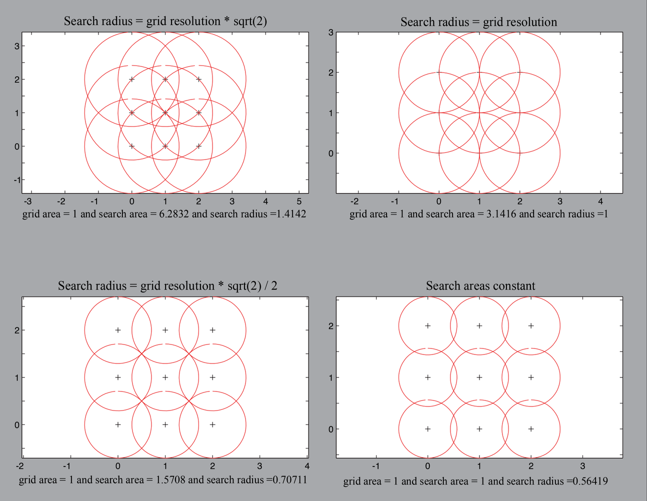
PROPOSED GEON-BASED MODEL FOR ACCESSING AND PROCESSING COMMUNITY LIDAR DATASETS:



IV. DEM GENERATION BY LOCAL BINNING ALGORITH

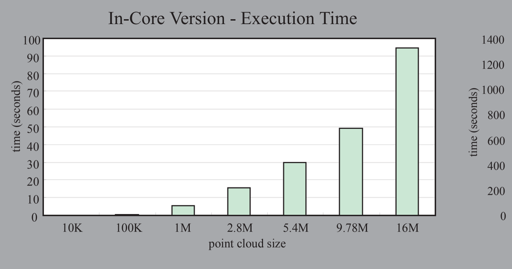
We have recently implemented a new local binning algorithm for DEM generation in the GLW.

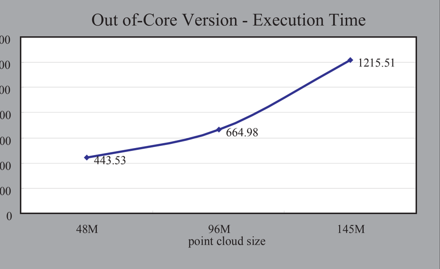
The algorithm utilizes the elevation information from the points inside of a circular search area with user specified radius. From the local points, five values are computed for each node in a grid: 1) the minimum, 2) maximum, 3) mean, and 4) inverse distance weighted mean of the local points, and 5) the number of points (density in the search area). If the number of points in the search radius is 0, the node is assigned a null value.



PERFORMANCE:

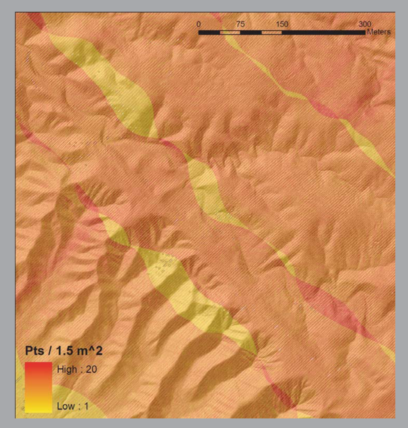
The GRASS spline algorithm implemented in the GLW is compute-intensive with large (millions of points) datasets taking tens of minutes to hours to complete. The local binning algorithm offers approximately 100 times better performance than the spline. For large grids that exceed main memory on the compute node, an out of-core version of the code utilizes secondary storage to process the data in segments. A parallel version of the out of-core code is under development and should be available via the GLW in the near future.



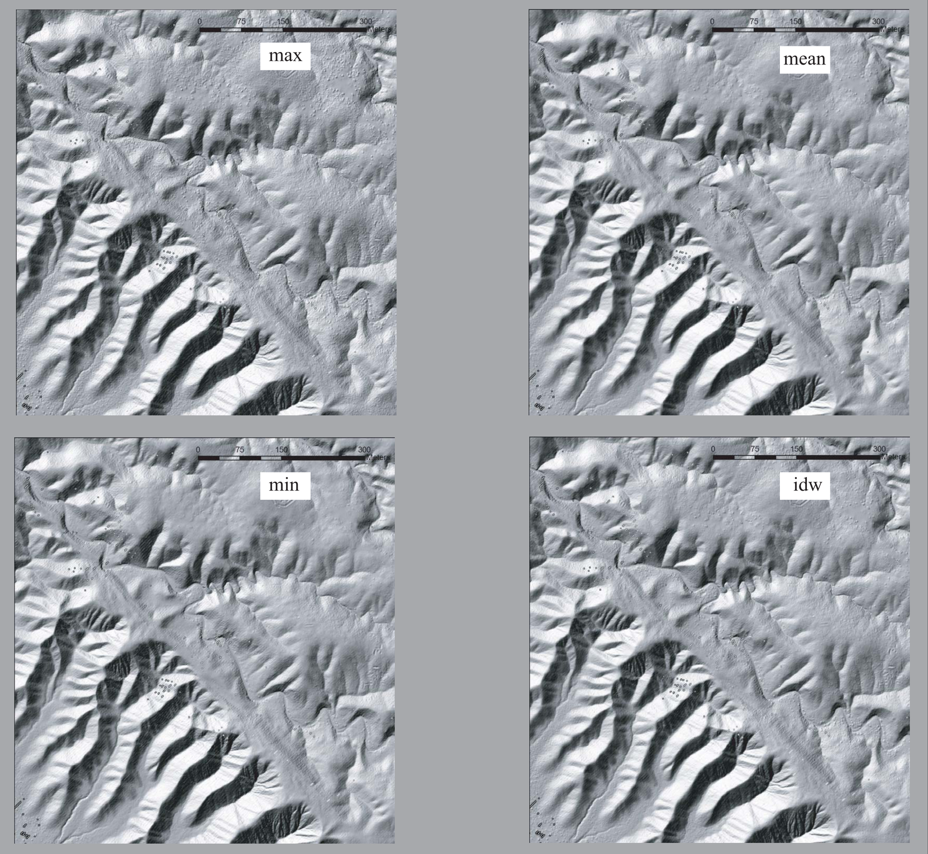


EXAMPLES:

RIGHT: Figure showing the number of LiDAR returns per search area for a portion of the B4 dataset in the Carizo Plain near the Dragon's Back. For this 1 m DEM, a search radius of .70711 m was used, giving a search area of 1.5708 m^2. Note the heterogeneous LiDAR return distribution due to swath overlaps and changes in plane orientation. LiDAR return concentraions vary from 1 to 20 per search area. This grid was produced from ~4.7 million LiDAR points.



BELOW: Four hillshaded 1m DEMs for the same area as shown at right. All four DEMs are produced by the local binning algorithm implemented in the GLW. The local binning approach works well when the DEM resolution is greater than the LiDAR return spacing. When the DEM resolution approachs the shot spacing, a large search radius or an interpolation algorithm (e.g. spline) must be used to avoid numerous null values.



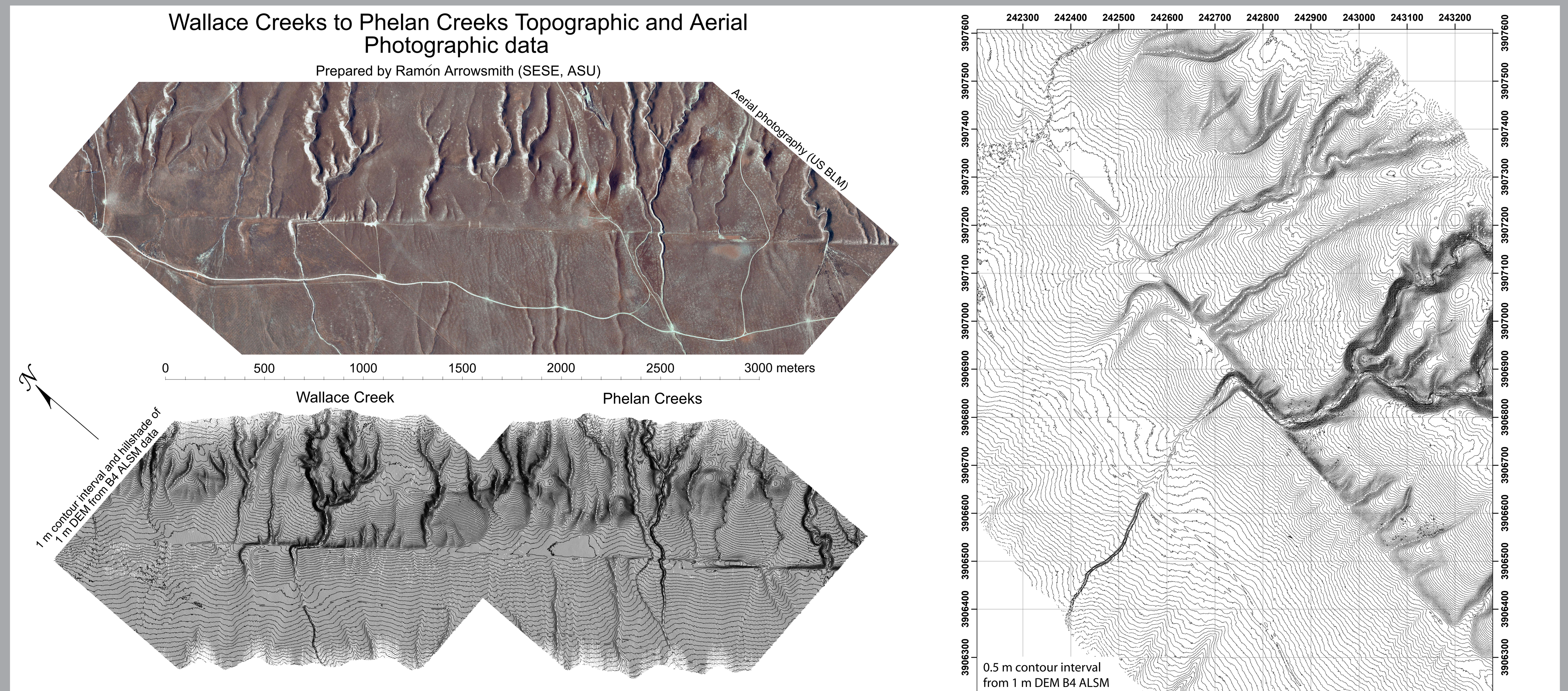
VI. STATUS & ACCESS INSTRUCTIONS

- Currently, the southern San Andreas portion (segments SAF1-11) of the B4 dataset is available via the GLW, with the Banning and San Jacinto segment to follow shortly.

MORE INFORMATION:
http://www.geongrid.org/science/lidar.html
http://lidar.asu.edu

The GLW can be directly accessed by logging into the GEON Portal @ **http://www.geongrid.org** (users must register for a free account), then selecting the "GEON Tools" tab and then "LiDAR".

V. APPLICATION—TECTONIC GEOMORPHOLOGY ALONG THE SAN ANDREAS FAULT

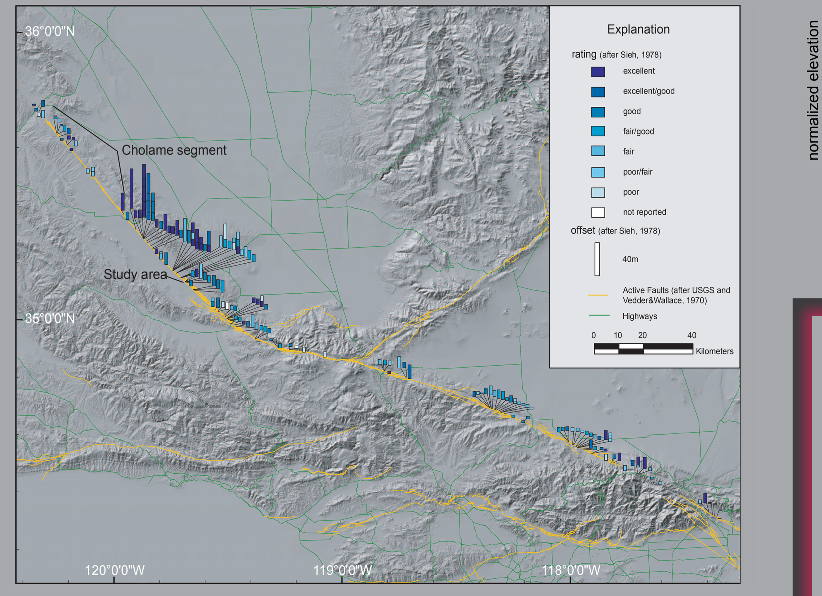


New versions of the classic depiction of the San Andreas Fault in the Carrizo Plain at Wallace Creek

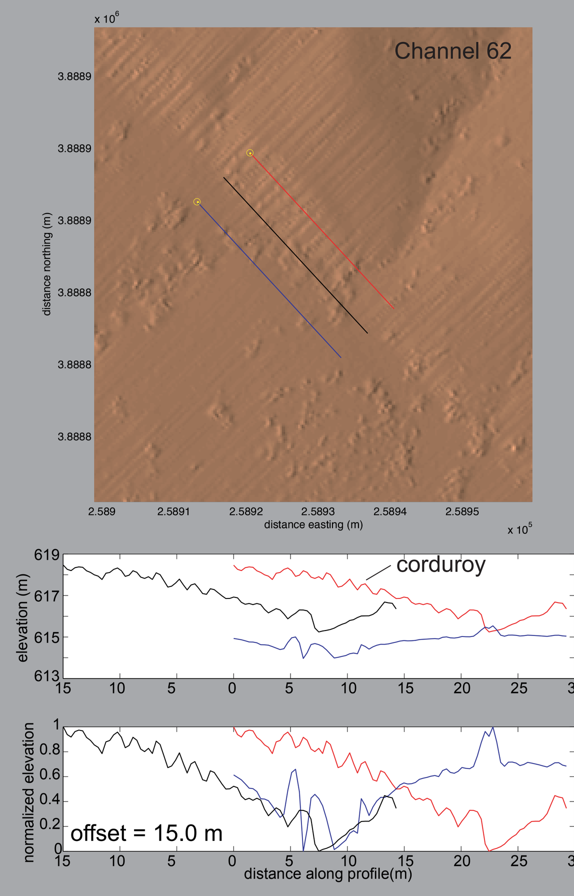
Problem: slip along the San Andreas Fault in 1857 earthquake and earlier

Offset stream channel along SAF in Carrizo Plain depicted in 4 cm resolution georectified aerial photo from remote control balloon system draped over 25 cm from B4 survey and GEOn Lidar workfow

Important information about paleoearthquakes is recorded in the meter-scale tectonic geomorphology of fault zones. Such information is available from both the records of offset and from paleoseismic sites which are better understood with enhanced knowledge of geomorphic context. The B4 airborne laser swath mapping dataset of the topography along the San Andreas Fault (SAF) zone has enabled a spectacular new look at the 10s-of-meter scale tectonic geomorphology and is a powerful complement to field survey.



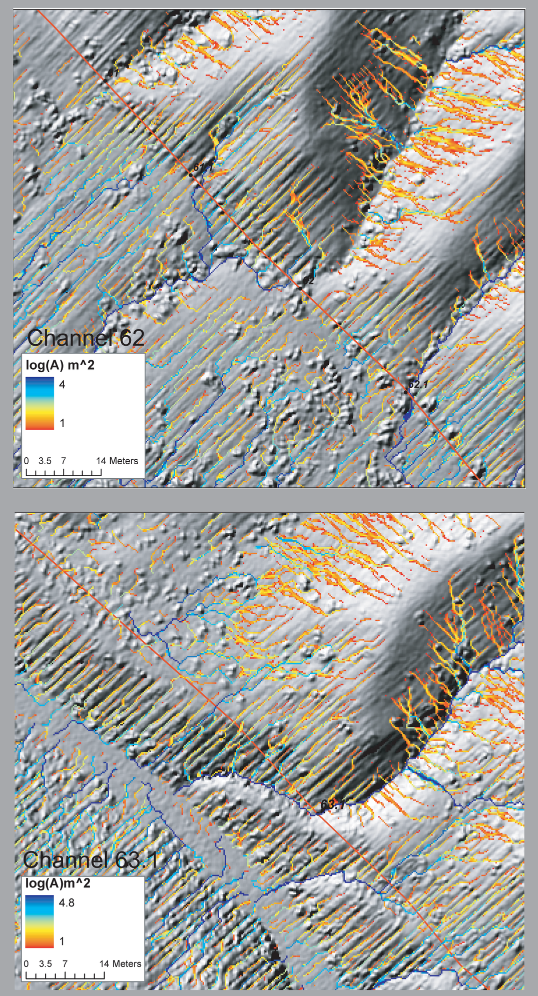
Offsets along the San Andreas Fault in Southern California (Sieh, 1978). The smallest offsets in this group are attributed to the 1857 earthquake. Offsets along the northwestern portion of the Cholame segment were also studied by Lienkaemper. The south central SAF preserves numerous offset landforms that will be uniformly charaterized in detail with high resolution ALSM. The entire SAF shown in this map has been imaged with ALSM. Active faults in yellow from USGS and Vedder and Wallace, 1970.



One basic idea of determining offsets is to assume that the channel was similar in profile on each side of the fault, so offset determination comes from simply matching parallel profiles. In this example, I took a 25 cm DEM from the GLW at channel 62 (see above) and extracted a profile 5 m on each side of the AF trace. The profiles correlate best with an offset of 15 m which is similar to the 15-17 m offsets determined above.

The problem is that this approach removes the geomorphological intuition that is typically employed as the observer ponders the subtle history of the channel: Was the channel perpendicular to the fault trace prior to offset? Has it been offset more than once? (Should have been in this case.) Where is the evidence for channel development between the first and second offset?

Note also the ALSM scanner parallel grooving ("corduroy") from superposition of mislocated swaths. It has an amplitude and wavelength of 20-50 cm.



Another way to have the DEMs help the tectonic geomorphologist is to produce hydrologically correct DEMs (fill pits) and then compute drainage network parameters, such as contributing area (A). Such information should help to guide the observer in semi-automatically delineating the channel network, and thus the offsets.

In this case, I show examples from channels 62 and 63.1 on 25 cm DEMs from the GLW

Channel 62 comments:
1) Note that the network delineation is disrupted by the vegetation in the channel--indicating the need for vegetation filtering.
2) I measured an offset of about 13 m, which is about 2 m less than the other approaches.

Channel 63 comments:
1) The channel is well defined in the hillshade and in the contributing area computation.
2) I measured an offset of 19 m which is in the middle of the 15-21 m measurements from the other methods.

This material is based upon work supported by the National Science Foundation under Grant No. 0225543 (GEON) & the Southern California Earthquake Center

REFERENCES:
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.

Acknowledgements
Olaf Zielke, Kelin Whipple, Kevin Eagar, David Schwartz, Nathan Toke, Sinan Akciz
B4 Team, NCALM
GLW team

Abstract
LiDAR (Light Distance And Ranging a.k.a. Airborne Laser Swath Mapping—if airborne and Terrestrial Laser Scanning—if ground-based) is an emerging technology for acquiring high-resolution digital terrain models. Typical airborne LIDAR surveys sample the ground surface and vegetative cover multiple times per m2 while ground-based LiDAR samples at centimeter densities. These data and the associated panchromatic, color, or hyperspectral imagery that may be collected simultaneously provide unprecedented representations of the earth's surface (topography and vegetation) enabling the study of natural processes at resolutions not previously possible yet essential for their appropriate representation. Community-oriented LiDAR acquisition efforts (e.g., GeoEarthscope and NEON) will produce datasets with billions of LiDAR returns and terabytes of imagery data. Managing, archiving, distributing, and processing such data represent a significant computational challenge that can limit the utility of these datasets for scientific users. Based on a prototype system developed in GEON, we propose a national cyberinfrastructure framework for community LiDAR datasets that would significantly ease access to these datasets by providing internet-based data distribution and processing tools. The architecture we propose capitalizes on a distributed network of data servers and processing resources that are accessed via a web portal. This distributed network utilizes modular web services and is seamlessly linked to allow users to interactively access and process regardless of where the data or processing tools are physically located. This framework allows users to select a subset of data, generate terrain models as well as derived products with processing parameters they define and to download and/or visualize their results depending on user specification. The national cyberinfrastructure framework for LiDAR presented here both significantly enhances the utility of these data for scientific users as well as provides a clear example for distributing and processing computationally challenging community earth science datasets. The GEON prototype (the GEON LiDAR Workflow) is currently available for use via the GEON Portal (portal.geongrid.org).