An Internet-based Tool for Accessing and Processing the Southern San Andreas (B4) LiDAR / ALSM Dataset

http://activetectonics.asu.edu
http://lidar.asu.edu

Christopher J. Crosby[1], J Ramón Arrowsmith[1], Efrat Frank[2], Viswanath Nandigam[2], Han Suk Kim[3], Jeffrey Conner[1], Ashraf Memon[2] and Chaitan Baru[2]

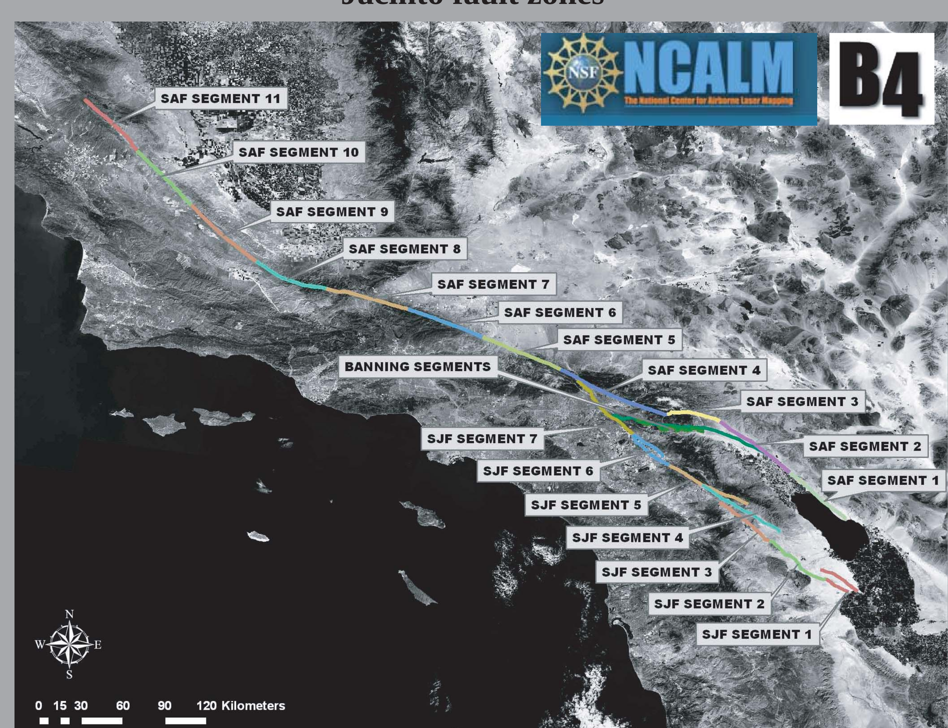
[1]*School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287*
[2]*San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093*
[3]*Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093*
chris.crosby@asu.edu

GEON CYBERINFRASTRUCTURE FOR THE GEOSCIENCES

I. INTRODUCTION & MOTIVATION

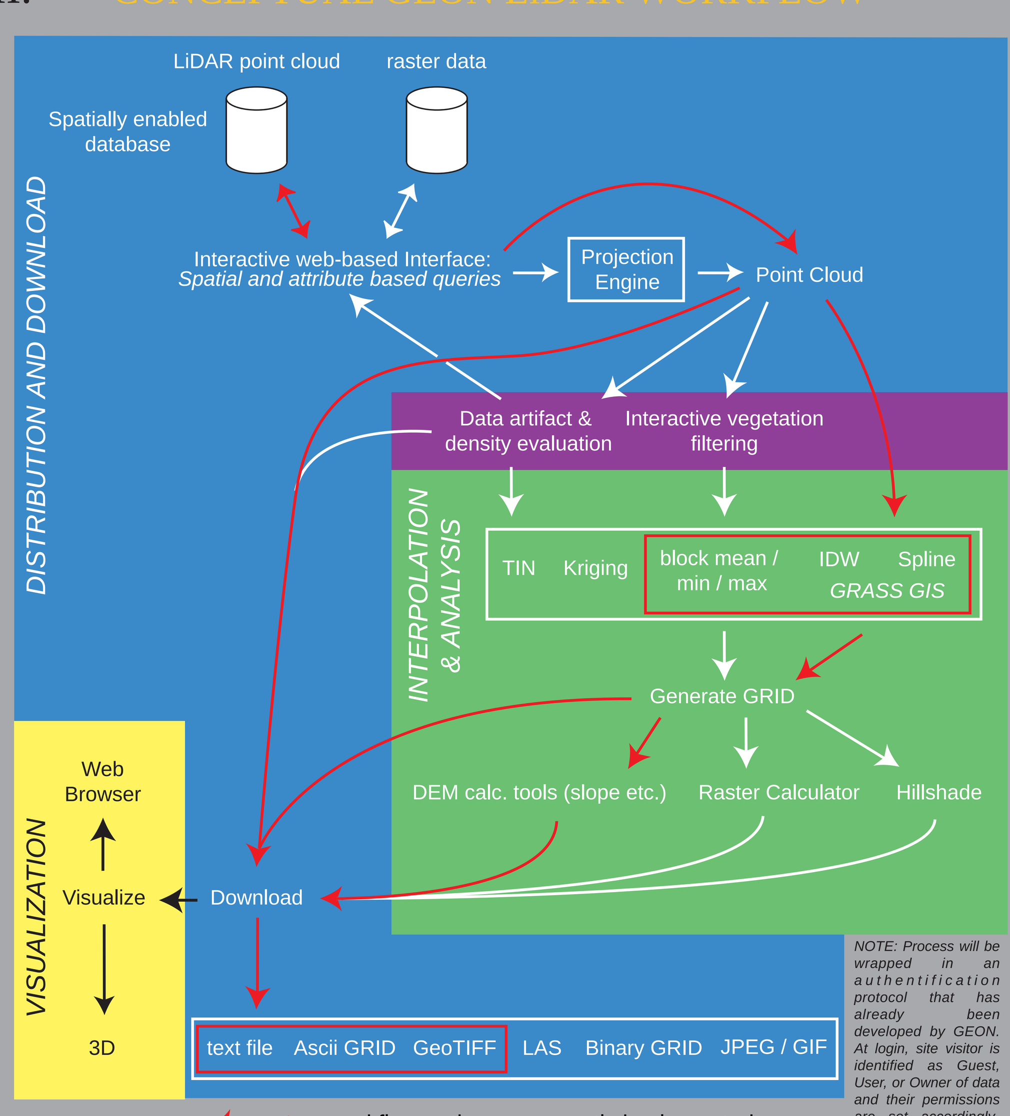
- LiDAR / ALSM generates massive data volumes - billions of returns are not uncommon in these data sets.
- Processing and analysis of these data requires significant computing resources not available to most geoscientists.
- Geoscience users typically work with digital elevation models (DEMs) generated 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: 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:

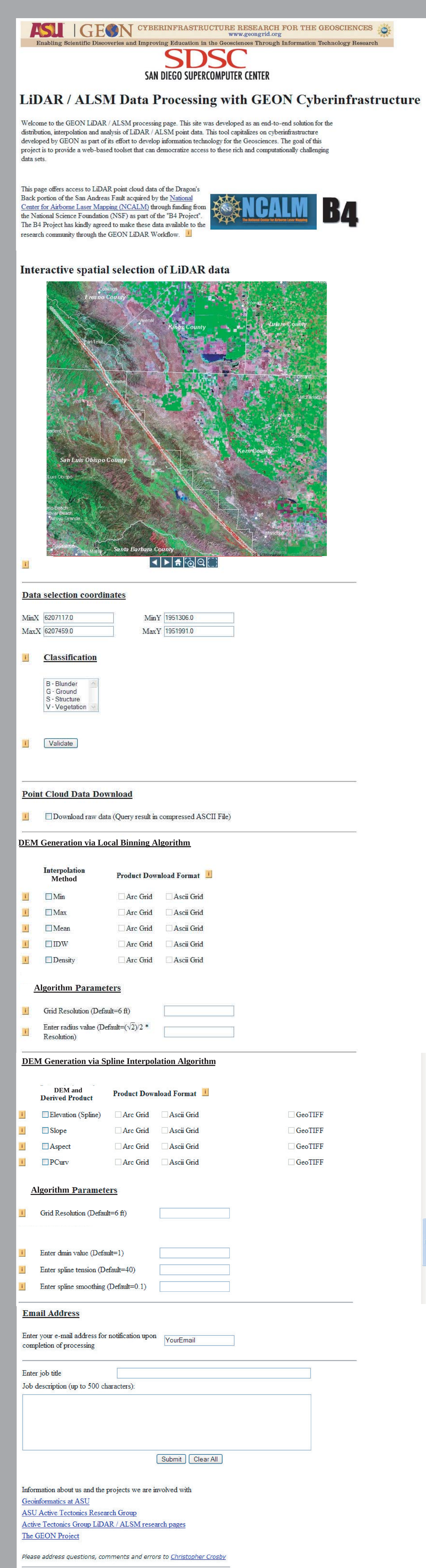
- Using GEON cyberinfrastructure we have developed an internet-based LiDAR distribution and processing (DEM and derived product generation).

Features:

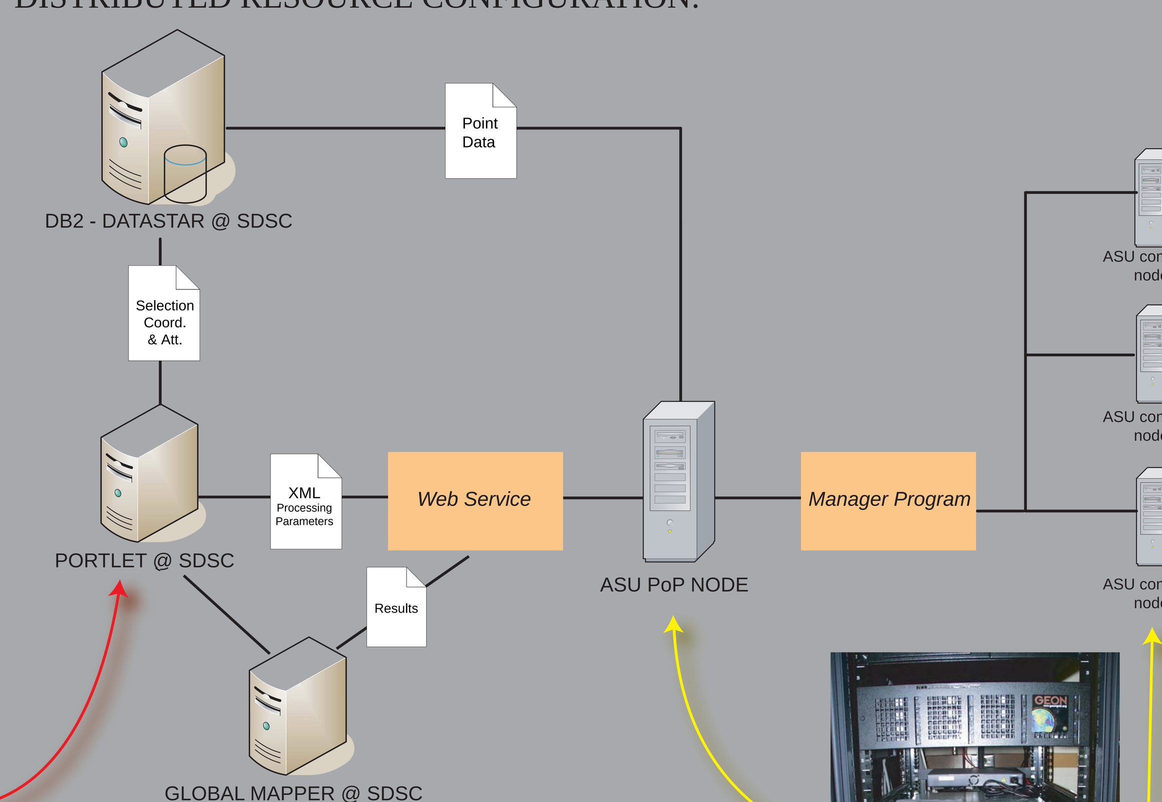
- Spatial and attribute based queries on raw LiDAR point cloud data.
- Local binning or Spline interpolation to Digital Elevation Model (DEM).
- User control over interpolation parameters.
- Slope, aspect and profile curvature (pcurv) derived products.
- Download of of products in TIFF (with world file), 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

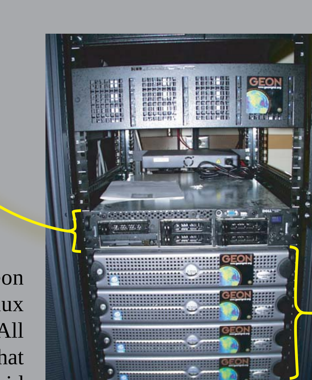
IMPLEMENTATION:

- 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)



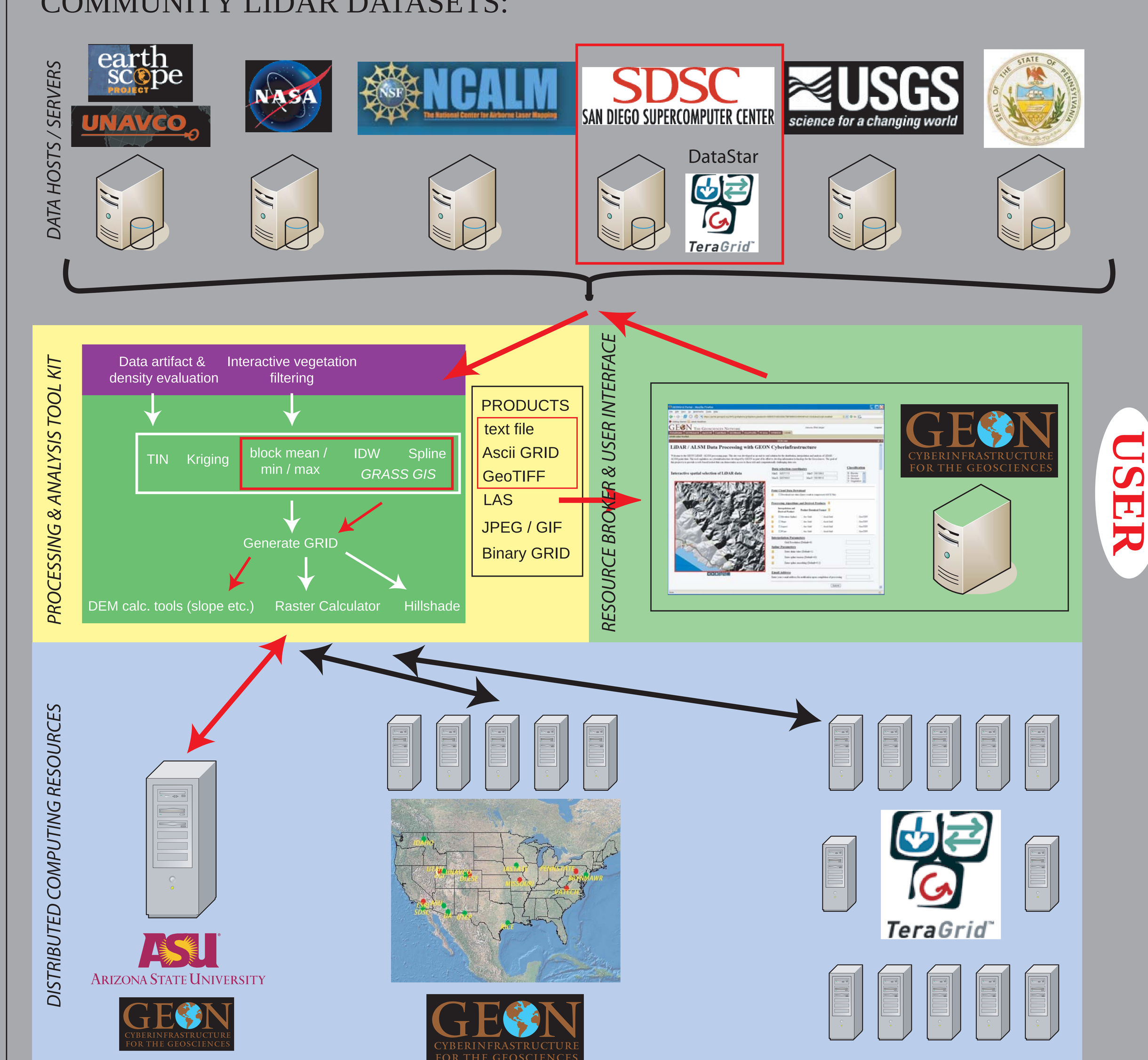
DISTRIBUTED RESOURCE CONFIGURATION:





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

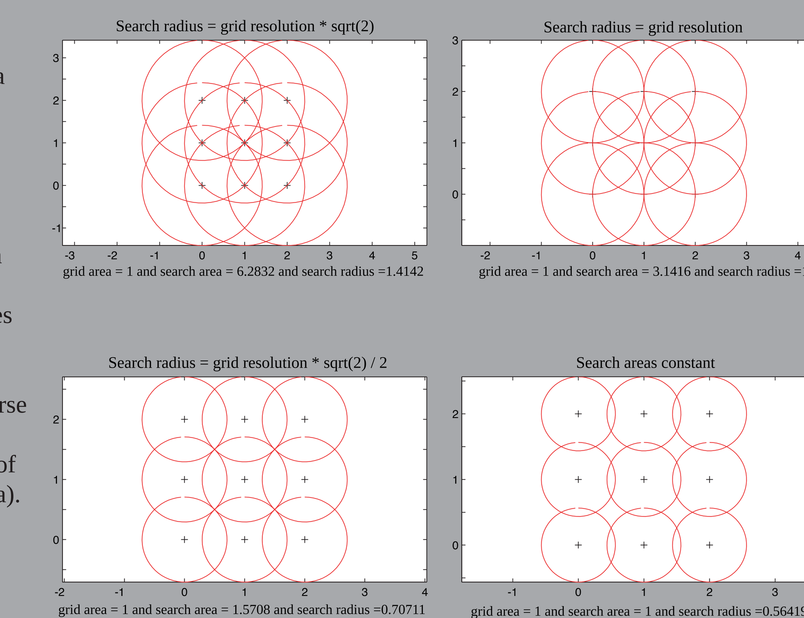
PROPOSED GEON-BASED MODEL FOR ACCESSING AND PROCESSING COMMUNITY LIDAR DATASETS:



IV. DEM GENERATION BY LOCAL BINNING ALGORITH

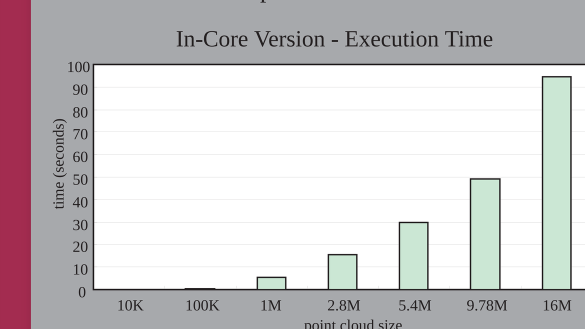
We have recently implemented a new local binning algorithm for DEM generation in the GLW.

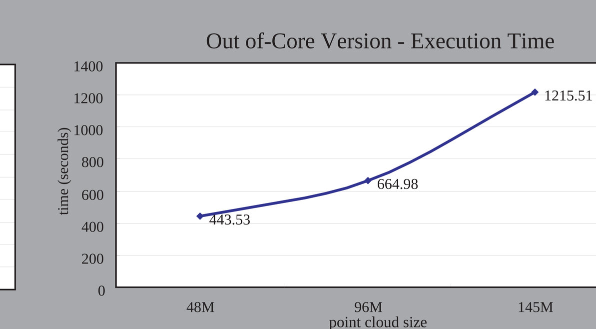
The algorithm utilizes the elevation information from the points inside of a circular search area with user specified radius. From the local points, five values are computed for each node in a grid: 1) the minimum, 2) maximum, 3) mean, and 4) inverse distance weighted mean of the local points, and 5)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

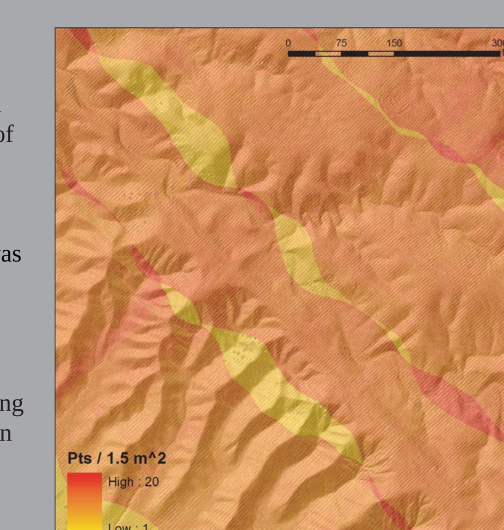
The GRASS spline algorithm implemented in the GLW is compute-intensive with large (millions of points) datasets taking tens of minutes to hours to complete. The local binning algorithm offers approximately 100 times better performance than the spline. For large grids that exceed main memory on the compute 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.



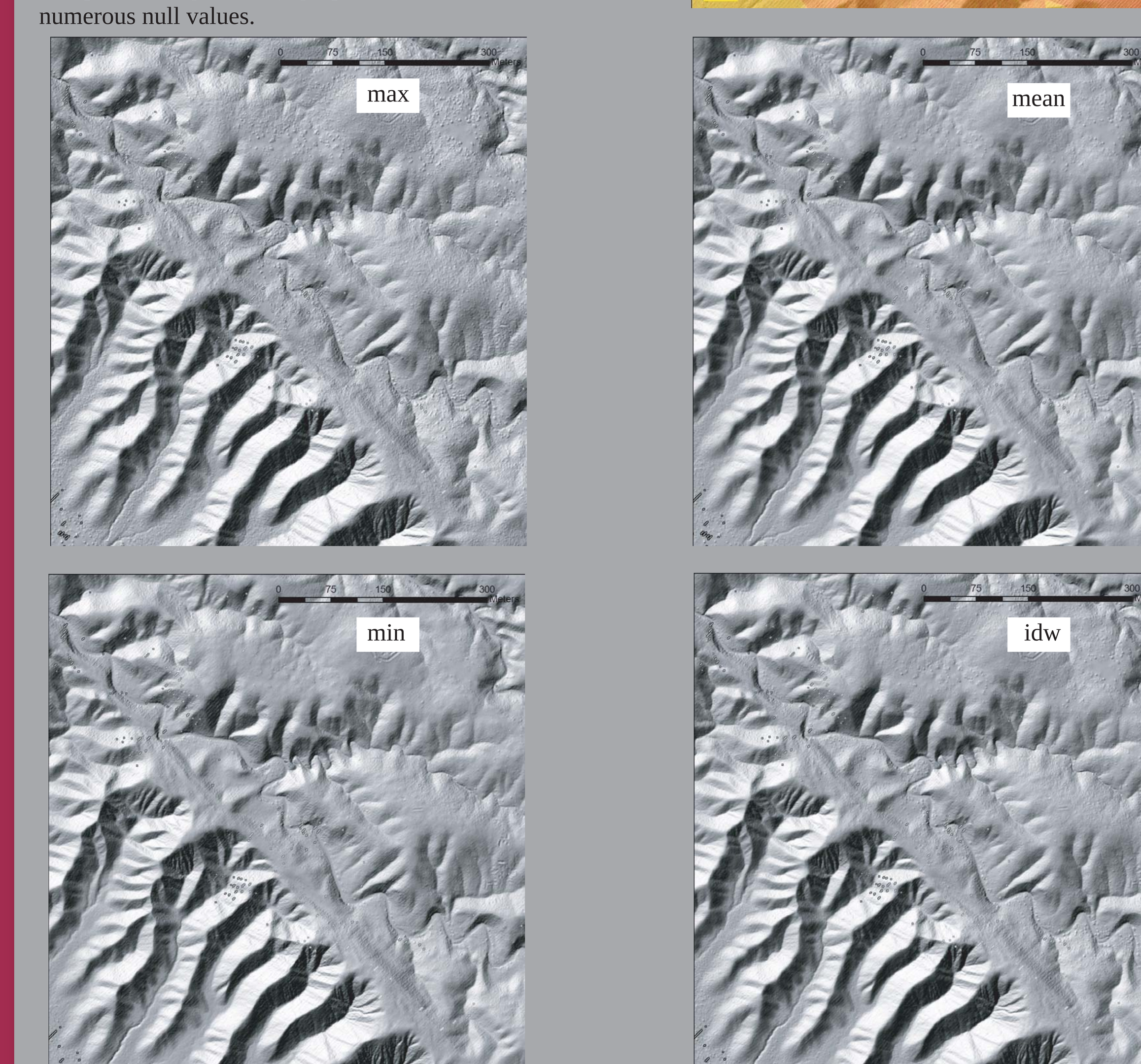


EXAMPLES:

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.



V. 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".

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.

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