**Solid Earth Science**

Some scientific areas have explosive drivers for the data deluge. We have seen this for LHC accelerator, gene sequencing, weather/climate, Astronomy and Exascale visualization. Drivers include both technology improvements and also large increases in deployment due to pervasive need of say personal health or weather data. Other fields see data deluge but at a lower magnitude. Here we discuss earthquake and polar science.

Fortunately for society if not science, large earthquakes are infrequent and so the study of earthquakes is data-limited compared to other fields. Major quakes occur all over the world and it is unrealistic to have substantial sensors deployed for most of these. Further, the quasi-periodicity of earthquakes implies that historical data is very important and we cannot increase that. Simulations can forecast damage and perhaps the aftershocks of an earthquake but the most important capability – forecasting new quakes is essentially entirely observational. Typically one uses patterns (in time series) to forecast the future with simulations useful to check if a particular pattern informatics approach is valid in an ensemble of simulated earthquakes. Important types of data include

1. Catalogs of Earthquakes with position and magnitude
2. Geometry of Earthquake faults
3. Global Positioning data (GPS) recording time dependent positions
4. Synthetic Aperture Radar inferograms InSAR recording changes in regions over time.

The first two pieces of data a) and b) are gathered carefully with recording of earthquakes and field analysis; this is small in size and growing slowly in size but of very high value. There are currently less than 10,000 GPS stations recording data at an interval varying between one second and a day. Well-known GPS networks are the Southern California Integrated GPS Network, the Bay Area Regional Deformation Network in Northern California and the PBO Plate Boundary Observatory from UNAVCO.

The inSAR data could become voluminous but currently totals some 350 images (each covering around 10,000 km2) and 2 terabytes in size. This data comes from uninhabited aerial vehicles (UAVSAR from JPL) or satellites (WInSAR from UNAVCO). The situation could be revolutionized by the approval of the DESDynI-R Mission (Deformation Ecosystem and Dynamics of Ice– Radar) recommended in the Earth Science Decadal Survey. DESDynI would produce around a terabyte of data per day but the mission is not approved and so is many years away from a possible launch. This data is analyzed (as by QuakeSim for recent earthquakes) to find rate of changes, which are then used in simulations that can lead to better understanding of fault structures and their slip rates.

*Figure 1: Architecture of PolarGrid data analysis Cyberinfrastructure*

Turning to Polar science, a good example here is the work of the CReSIS (Center for Remote Sensing of Ice Sheets) led by Kansas University that is pioneering new radar and UAV’s to be used to study ice-sheets. Multiple expeditions fly instruments that collect data including: (1) ice thickness and internal layering from radar and seismics, and Synthetic Aperture Radar (SAR) images of ice-bed interface; (2) bed topography generated from ice thickness and surface elevation); (3) time series of change in surface elevation from airborne and satellite altimeters; (4) time series of surface velocity from repeat-pass satellite images, in situ GPS measurements, and aerial photos; and (5) bed characteristics such as temperature, wetness, and sediment from seismics and radar. The last CReSIS expedition took 80 terabytes of data in 2 months. After traditional processing with FFT’s, radar images are produced along multiple flight lines as illustrated in figure 2. Then image processing is needed to identify the top (red) and bottom (green) of an ice-sheet. Initially students performed this but recently it has been automated with an image analysis tool developed by David Crandall at Indiana University. The deployment of UAV’s rather than current Orion and DC-8 conventional aircraft will increase data gathering capability by allowing continuous operation. There are more complex data such as snow deposits showing annual layers revealing historical snow deposition.



*Figure 2: Radar imagery from CReSIS*

The sea-bed data illustrated in figure 2 is fed into simulations that aim to understand the effect of climate change on glaciers. Note that gathering of data is complicated by the paucity of electrical power and poor internet connectivity to the polar regions. GPU’s are interesting as a possible lower power processing approach. The data is gathered on removable disks mounted in a storage array connected to just one or servers with rugged laptops as personal machines. All software is written in Matlab.

**References**

DESDynI <http://desdyni.jpl.nasa.gov/mission/>

QuakeSim <http://www.quakesim.org/>

UNAVCO and WInSAR <http://www.unavco.org/>

Alaska Satellite Facility <http://www.asf.alaska.edu/>

UAVSAR <http://uavsar.jpl.nasa.gov/>

CReSIS <https://www.cresis.ku.edu/>