Geospatial and Global Earth Mapping and Modelling Applications Requirements for the BDEC2 Workshop

William Kramer[1], University of Illinois at Urbana Champaign
with contributions from Paul Morin, Jonathan Pundsack, Claire Porter and others at the Polar Geospatial Center[2] at the University of Minnesota, and Brett Bode and Greg Bauer, NCSA[3] at University of Illinois at Urbana Champaign
November 28-30, 2018

This white paper is an initial and limited discussion of requirements and goals for geospatial information systems and applications. GIS is broad umbrella for a range of research uses from mapping and imaging the earth – not just the surface – to providing assessments of current and past conditions augmented with strong modeling capabilities, including the integration of physical models with social and environmental models. Geospatial information already integrates expensive sensors (e.g. high resolution satellites) with many, lower cost sensors (field level, UAVs, radar, lidar, sonar, etc.) at all scales, which all produce huge amounts of raw data and then uses various computational and analysis methods (image analysis, computer vision, model creation, ML/AI, modeling and simulation, …) to derive insights.

The scope of geospatial activities is too large and varied to incorporate into a single, short white paper, so the remaining paper will focus on the needs, requirements and impacts of just one fundamental enabling aspect for GIS - the new method of producing high resolution Digital Elevation Models (DEMs) from satellite images. In the last few years, new best of breed approaches created paradigm shifting map creation capabilities that greatly reduce the cost and improves the timeliness and resolution of traditional map making. The improvements include increasing the resolution of current elevation maps by more than 3 orders of magnitude (12.5^2 improvements in resolution), improving the time to production 58,500x in time to solution compared to a single workstation, and a 220 times reduction in cost, resulting in 9 orders of magnitude of overall productivity improvement. The proof of the paradigm shift has been shown in the ArcticDEM[4,5] and Reference Elevation Model of Antarctica[6] (REMA) projects that use the NCSA's Blue Waters NSF leadership computing system to generate essentially complete elevation maps using Illinois Innovation and NSF PRAC allocations. The improvements mean it is for the first time feasible to envision creating very accurate Digital Elevation Models on a global and frequent basis.

To give some idea of fundamental, albeit limited, requirements for computing and data required for creating two meter accurate DEMs of the entire landmass of the earth. The new five year EarthDEM project extends the impact and scope of these methods to the entire land mass of the earth.

Global DEM processing requirements

The goal of EarthDEM is use a global coverage set of in-track and cross-track satellite images to create two meter digital elevation models of the entire landmass of the earth and provide those to the public in mosaics and well as localized data sets.

[1] All error and mistakes in this document are the sole responsibility of the author and not the contributors
[2] https://www.pgc.umn.edu/data/arcticdem/
[3] http://www.ncsa.illinois.edu
[4] http://www.ncsa.illinois.edu/news/story/blue_waters_processes_final_installment_of_arcticdem_mapping_initiative
[5] https://bluewaters.ncsa.illinois.edu/liferay-content/document-library/18symposium-slides/porter.pdf
[6] http://www.ncsa.illinois.edu/news/story/ncsas_blue_waters_supercomputer_helps_map_the_poles

	Strips	Area – km^2	Estimated BW node-hours required to process one time	Data Required (PB)
In-track	336,492	538 million	54 million	5.38
Cross-Track	2,956,131	2,365 million	473 million	47.30
Total	3,292,623	2,903 million	527 million	52.68

Based on Polar Geospatial Center (PCG) staff estimates, the computational needs to create a single, global set of DEMs once is 527 million Blue Waters X86 node hours[7] that are needed to process all the EarthDEM Digital Globe/NGA imagery that could be provided. The Arctic and Antarctic areas represent about 1/6 of the earth's land mass surface area each. Given the need to do reprocessing for areas (e.g. if clouds are present in the first set of images) and other factors, the total processing of all the data is approximately 3.2-3.5 dedicated, sustained "Blue Waters Years" of computing.

For data requirements, each stripe averages 4 GB in size, and two strips are needed for every DEM. Two meter DEMs average 8 GBs. So, the processes consumes 8 GB and produces 8 GB per sample. This indicates in addition to over 527 billion node hours of computational time, the processing requires 26-30 PB is consumed and 26-30PB is produced. Since the original strip data flows from repositories specific for the satellites, and is stored in open access repositories, these 50+ PBs of data has to move within the period of the campaign. If you assume this is a yearly campaign, the average sustained data rates are ~10 kbps, but will have peaks where multiple streams of 8 GBs need to move before or after a job initiated.

Benefits and Impacts of DEM creation

Having accurate DEMs is a necessary condition for many other research and practical areas, all of which rely on additional processing and analysis to gain their final insights. For example, for hydrology, DEMs are used to develop very large graphs that can be used to analyze water flows and storage, flooding and drought potential, water use and watershed contamination. For weather and climate, accurate DEMs can be used for analyzing micro climate and micro weather predictions and well as enhancing the predictability of severe weather based on the land surface atmosphere interaction. High resolution DEMs are used for understanding earth surface changes, such as impacts of mud slides and earthquakes, analysis and potentially predication of volcano eruption, etc. DEMs can be used for natural resource management such as forest and agricultural management and wild life migration patterns. DEMs also can be used for analyzing ice melts, glacier retreats and advances, sea level rise, etc. all on a larger, much more timely manner and lower cost scale than previously possible. Urban analysis and planning also now can use high resolution DEMs since at, 2 meters, (and even 30 cm when high accurate reference elevations are present) clearly show buildings, roads, infrastructure, etc.

Computational and data analysis resources need to support all these derived uses and more areas that are enabled by the use of high resolution DEMs. Just as DEMs are now much more accurate, the derived investigations will want to do more accurate and detailed analysis, which will require significant increases in computational and data resources. Furthermore, to facilitate use of DEMs and other data products by others, effective infrastructure is needed to enable easier

[7] A serious attempt was made to port SETSM to the GPU based XK nodes in the BW PAID program, but the resulting code ran slower than the XE version. Further work would be needed to see if SETSM could benefit from GPU acceleration.

storage, annotation and retrieval of the original images, the DEMs, metadata and derived data products.

Frequency of DEM releases

ArcticDEM has made two major, complete releases of their mosaic DEM, the first after about 18 months of processing on Blue Waters and a second after about three years of processing. Antarctic REMA has made one data release at 8m after about 18-24 months of processing. The geospatial community who use DEMs desire complete global DEMs every couple of years, and more frequent DEMs in areas of high interest such as watersheds, areas of earth crust movement such as the Pacific "ring of fire, etc. maybe on monthly basis or even on demand basis. In order to do this on a regular and reliable basis will require a substantial fraction of exascale computing resources.

Other Geospatial BDEC activities

Of course, DEMs and their use are only one part of the geospatial research and production, many of which have similar levels of requirements. Many other methods are in use and will need to be supported on BDEC resources. Some of these areas are image classification, identification and isolation, crop and vegetation analysis, computer vision, weather and climate, oceanography, human population responses to events be they droughts and floods, to infectious disease outbreaks and other stimuli. All these methods require large amounts data from many sources and very significant amounts of computation and data analysis.