

The Virtual Exploration Geophysics Laboratory (VEGL) portal: Harnessing cloud computing for potential-field inversions

Richard Chopping*

Deep Exploration Technologies CRC
Geoscience Australia
GPO Box 378 Canberra ACT 2601
richard.chopping@ga.gov.au

James Goodwin

Geoscience Australia
GPO Box 378 Canberra ACT 2601
james.goodwin@ga.gov.au

Shane Bailie

CSIRO
Building 1A Wilf Crane Cres Yarralumla ACT 2600
Shane.Bailie@csiro.au

Josh Vote

CSIRO
26 Dick Perry Ave Kensington WA 6151
Josh.Vote@csiro.au

Ben Evans & the NCI staff

National Computational Infrastructure (NCI)
Leonard Huxley Building (#56), The Australian National University
Canberra ACT 0200
Ben.Evans@anu.edu.au

SUMMARY

Geophysical modelling often requires High-Performance Computing (HPC) when applying complex calculations to large datasets. To meet HPC requirements, access to parallel grid computing services via the internet, termed cloud computing, has been established for geophysical inverse modelling using UBC-GIF codes GRAV3D and MAG3D. Portals and workflows have been developed to simplify the inversion modelling process and to allow more efficient access to inversion modelling tools.

A mechanism which allows inversion modelling jobs to be submitted to cloud computing resources using a web browser (a geophysical portal) is being developed and tested at the National Computational Infrastructure (NCI) housed at the Australian National University (ANU). This work is the result of a collaborative project between Geoscience Australia, CSIRO and NCI.

Key words: potential-field inversions, cloud computing

INTRODUCTION

Geophysical modelling, especially 3D inverse modelling, requires significant computational power due to the application of complex calculations to large datasets. The need for HPC in the field of geophysics has led to the implementation of parallelised computing codes (e.g. Goodwin *et al.*, 2012, this volume). In addition, cloud computing is another avenue that can provide access to HPC resources. Cloud computing, supported by providers such as Google, Microsoft, Amazon and others, refers to a multitude of technologies. Here, we refer to cloud computing as access to distributed or grid computing where a software stack, suitable for geophysical modelling, is provided. In particular, access to the University of British Columbia-Geophysical Inversion Facility (UBC-GIF) inverse modelling tools GRAV3D (Li and Oldenburg, 1998) and MAG3D (Li and Oldenburg, 1996) is included in the software stack.

Below describes a test case where both a suitable interface and workflow have been established to enable GRAV3D and MAG3D inversions to be performed in a cloud computing environment.

GRAV3D / MAG3D WORKFLOW

To successfully enable GRAV3D and MAG3D inversions to be performed using cloud computing, a suitable workflow needs to be developed. Ultimately, there are multiple ways of preparing and processing data for use in GRAV3D and MAG3D inversions and, as a result, any workflow needs to be flexible enough to account for these differences. The common operations performed, and some optional operations, on input data are shown below in Figure 1.

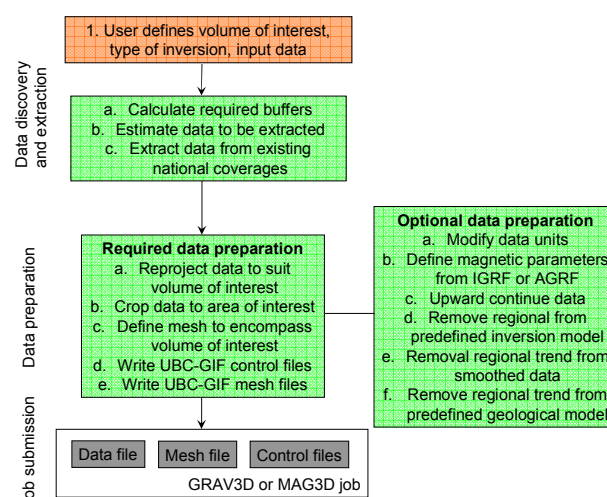


Figure 1. Simplified workflow for creating a GRAV3D or MAG3D inversion. A user must first define their volume of interest and desired model projection – this defines the required data to be extracted. These data must be further manipulated to be used as inputs to the inverse modelling routines. A number of the steps, both required and optional, are universal between gravity and magnetic inversions.

A number of the steps taken in the inversion, e.g. selecting buffer sizes or required data to be extracted, can be automated using rules of thumbs (such as formulae presented in Williams et al., 2008). Other steps such as projection of data and transforming to UBC-GIF formats are also easily scripted.

VEGL PORTAL

Both data discovery and scientific workflow portals (such as <http://portal.auscope.org/portal/gmap.html>) were developed within AuScope (www.auscope.org) which were subsequently used in the development of VEGL. An existing portal was modified to discover relevant geophysical datasets and to encapsulate suitable modules for manipulating data required for the potential-field inversions.

The test environment provided by the NCI is an implementation of OpenStack (www.openstack.org) and completed job files are currently stored using Amazon Simple Storage Service (Amazon S3: <http://aws.amazon.com/s3/>). The test implementation of VEGL has access to 8 nodes, each node containing ~12 GB of RAM and ~2.8GHz CPUs, running Linux. Cloud computing infrastructure, however, means that additional compute resources could be applied to VEGL to increase user capacity beyond a level suitable for testing, or to perform inversions faster.

There are four stages to using the VEGL portal (Figure 2):

1. Data discovery and extraction;
2. User access controls and data preparation;
3. Job submission and monitoring;
4. Automated metadata capture.

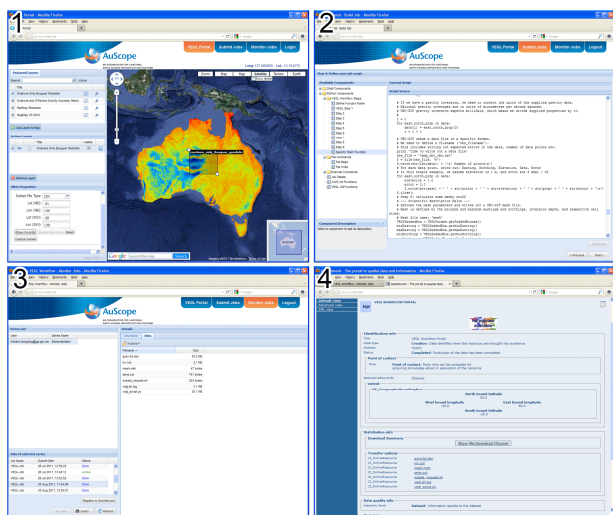


Figure 2. The Virtual Exploration Geophysical Laboratory (VEGL). In step 1, the user selects from available data and defines an approximate area of interest. This sets up a data extraction request from a web service. In step 2, the user is authenticated and develops a preprocessing script from pre-prepared Python code, as well as potentially developing new Python code. The data request and the preprocessing script are then submitted as a job in step 3. Finally, in step 4, metadata for the inversion job can be submitted to GeoNetwork – all required information to reproduce the inversion job is captured in this metadata.

In *Data discovery and extraction*, the user selects an area and required data. The data are discovered automatically and drawn on the fly on top of a Google maps interface. The data capture accounts for data projection and other inversion parameters. This replaces manual discovery and extraction and allows the user to obtain the most up-to-date data.

Next, in the *User authentication and data preparation* step, the user is first authenticated and then can select from pre-prepared scripts from a repository. These scripts can be edited as required by the science process. As errors are found in pre-prepared scripts or new methods of data preparation are developed, they can be incorporated and captured by the repository. As there is currently no consistency of data preparation for these types of inversions, this process greatly improves quality and consistency.

The use of pre-prepared scripts allows for significant flexibility while not requiring users to have high levels of programming skills. The scripts are written in Python and take the form of individual selected modules; in this way, optional operations such as particular types of filtering, can be performed as well as required steps such as formatting data for GRAV3D/MAG3D or writing meshes in the UBC-GIF format.

Within the data preparation step, users also have the ability to upload additional information, such as geological constraints or additional local-scale data. Any code developed within VEGL to incorporate such data can be captured and retained, and used by other geophysicists in their own future inverse models.

The third step, *Job submission and monitoring*, is where, following on from the data preparation step, the job is then submitted to the cloud to be computed. Computation of the job is performed by spawning a Linux virtual machine which then retrieves the appropriate data from a web service, runs the preprocessing Python script, and then executes the UBC-GIF inversion codes.

A user can monitor jobs to see their current states. Jobs are grouped by individual job series and by user. The job series is a useful method to keep track of inversions for different purposes, e.g. a geophysicist may be working in multiple projects and can group their work as such. Once a job has been completed, the results are uploaded to Amazon S3 storage to be retrieved later by suitably authenticated users.

Finally, the metadata of the inversion are captured in the Automated metadata capture step. In the third step, a button allows for the automated, one-touch capturing of all of the metadata associated with the inversion. This includes all files required to recreate the inversion, the script as utilised by the VEGL portal. In the test implementation of VEGL, metadata are captured to an implementation of GeoNetwork (<http://geonetwork-opensource.org/>).

BENEFITS OF VEGL AND CONCLUSIONS

There are significant benefits to using a portal such as VEGL to provide access to geophysical modelling codes such as GRAV3D or MAG3D. Firstly, the design of the data preparation step encourages reuse of modules. These modules can be updated if errors are found or new methods of data preparation are developed. Reuse also ensures consistency of modelling results. Secondly, the design of the portal lets the user focus on the scientific questions they are answering, rather than worrying about nuances such as recommended data buffering limits or other such factors. Thirdly, as the jobs are repeatable, models can be quickly repeated to test parameters such as matrix compression or other user-modifiable inversion parameters. Inversions may also be repeated when new data are made available. Users also are not restricted to those with the skills and/or software to manipulate data to the required formats. Inversions can be submitted by geophysicists with access to the internet and a web-browser, rather than requiring a high-end computer to process data and perform the inversions. Geophysical knowledge, however, is still required to ensure that inversions performed produce geophysically sensible results.

ACKNOWLEDGMENTS

A significant number of workers have been involved in the development of VEGL and due to space are not included on the authorship of this paper. In particular, many technical

support staff at the NCI have been crucial to the development of VEGL. Dr Lesley Wyborn (GA) and Dr Robert Woodcock (CSIRO) were also instrumental in this project, conducted under the GA-CSIRO eResearch collaboration. GA authors publish with the permission of the CEO, GA.

REFERENCES

- Goodwin, J., Williams, N.C. and Oldenburg, D.W., 2012. High-resolution regional-scale 3D inversion modelling using the National Computational Infrastructure. Proceedings of the 22nd International Geophysical Conference and Exhibition, 26-29 February 2012 - Brisbane, Australia (this volume).
- Li, Y. and Oldenburg, D.W., 1996. 3-D inversion of magnetic data. *Geophys.*, 61, 394-408.
- Li, Y. and Oldenburg, D.W., 1998. 3-D inversion of gravity data. *Geophys.*, 63, 109-119.
- Williams, N.C., 2008. Geologically-constrained UBC-GIF gravity and magnetic inversions with examples from the Agnew-Wiluna Greenstone Belt, Western Australia. PhD Thesis, University of British Columbia. http://www.eos.ubc.ca/ubcgif/pubs/theses/williams_phd_thesis.pdf