Scalable platform for geophysical data analysis and visualisation

P. Golodoniuc [1], A. Davis [2], S. Bradley [3], S. Mulè [4]

[1] CSIRO Mineral Resources, Kensington WA, Australia, pavel.golodoniuc@csiro.au
[2] CSIRO Mineral Resources, Kensington WA, Australia, aaron.davis@csiro.au
[3] CSIRO Mineral Resources, Kensington WA, Australia, sam.bradley@csiro.au
4 CSIRO Mineral Resources, Kensington WA, Australia, shane.mule@csiro.au

SUMMARY

Geophysical survey datasets used for resource exploration and detection are large in volume, dense in time and space, and have many dimensions. We present the Geophysics Processing Toolkit (GPT), an application for processing geophysical survey data prior to interpretation and inversion. Initially developed as a processing toolkit for airborne electromagnetic (AEM) data, our application can be extended to beyond electromagnetics processing and inversion to incorporate multiple geophysical datasets such as gravity and magnetics. Interactive visualisation and signal processing tools make the process more efficient. We have developed the GPT using a cross-platform technology stack designed to work in a containerized environment in a Cloud and accessible via a web browser. This approach makes it intrinsically scalable and cost-efficient to operate. The toolkit architecture allows for a greater degree of extensibility offering a range of interactive visualisations and integration of a suite of signal processing tools for noise detection and removal. Modern visualisation technologies allow the software to run on a standard workstation or a laptop while efficiently delegating all computationally intensive tasks to the accompanying Cloud-based processing unit. Decoupling of visualisation components from cloud compute and storage nodes allows on-the-fly substitution of analytical codes, e.g., forward modelling, inversions. This brings greater flexibility in experimental research through the ability to apply various numerical methods and compare results via elaborate visualisations and through the application of statistical methods. Delegation of computing tasks and storage requirements to a third-party cloud provider (a) minimizes procurement and maintenance costs of computing/storage infrastructure and (b) eliminates clients' privacy concerns as data are stored and processed in an isolated cloud environment.

SCOPE

To facilitate the development of new exploration methods we have required an operational platform that would provide elaborate interactive visualisation of large AEM surveys and flexibility to experiment with numerical codes without specialist knowledge in software application development. Interactive visualisations are at the core of data pre-processing — offering a suite of tools to inspect and remove noisy or inferior data before it can be further processed. Our approach to visualisation of the AEM survey data included the development of a dashboard that incorporates the survey plan (i.e. a map view) with topography, infrastructure, and satellite imagery along with the ability to visualise the decay profile for transients from individual soundings; and as individual soundings of decay-value versus time for single stations. These data views are linked together so that information in all three views can be updated and co-located. This assists in the quality assessment of AEM data before they can be inverted or transformed using a variety of methods and the results brought back for further visual analysis. Great attention is given to provenance. Recording all processing steps allows improved understanding of results leading to the reproducibility of complex workflows applied to complex data sets. This is particularly important since a significant amount of AEM data exists in the public domain and to improve the life cycle of the AEM data, we need to be able to maintain the digital value chain. Although computer programs with such functionality exist, the workflow for accessing the existing codes is often complex, slow and non-transparent, numerical codes are embedded and integration with new experimental codes is nearly impossible. Our solution aims to streamline the processing environment so that is capable of efficiently handling spatially and temporally dense datasets on the order of several gigabytes.

INTERACTIVE DASHBOARDS

Representing multi-dimensional data in static visualisations is a notoriously difficult exercise. Multi-view representation offers a variety of perspectives on the same information depending on the current context and user actions. To improve human-computer interaction a range of techniques are applied including interactivity, data scaling, data-focusing and linking (Buja A. et al., 1991; Buja A. et al., 1996). Observation of user behaviour and analysis of visual data perception has contributed to the design of visualisation components. Interactivity encourages engagement with the data and empowers users to explore and analyse large datasets from a range of scales and perspectives. Figure 1 demonstrates the AEM

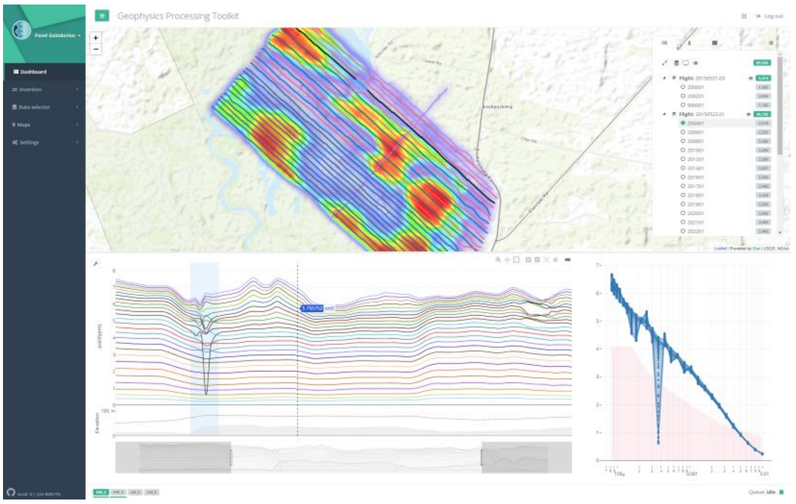

Figure 1: Interactive multi-view dashboard to provide interconnected multi-dimensional visualisation of the AEM survey data.

survey visualisation multi-view dashboard that provides an interconnected display of the AEM survey where actions in one view are reflected in all other views live. The dashboard provides functionality to overlay auxiliary GIS data and computer-generated visualisations, e.g., heatmap plots to show generalised representation and variability of geophysical properties.

SCIENTIFIC WORKFLOW

The GPT is a medium designed to connect users, visualisation tools, numerical algorithms, and computing infrastructure. Its architecture offers a range of extensibility interfaces – custom visualisation dashboards, data overlay widgets, numerical algorithms, as well as flexibility in the choice of cloud compute and storage infrastructures. We implement various visualisation widgets, such as, a generalised view of AEM survey geophysical properties via an interactive heatmap plot. Near real-time performance is achieved with a range of techniques, e.g., dimensionality reduction, use of pre-computed data, caching, etc. A suite of assistive technology algorithms was designed to simplify and guide the user through data quality control procedures, including automatic noise detection and removal. We implemented a proof-of-concept algorithm that detects variation in the rate of change between channels and highlights these areas on interactive charts. Algorithm sensitivity can be adjusted using a set of parameters to adapt to varying levels of amplitude in the input signal. A multitude of mathematical, statistical and geospatial algorithms can be applied to further improve its QA/QC capabilities. Toolkit architecture and cloud deployment strategy allow users to introduce experimental codes via microservices or externally hosted applications without compromising the intellectual property sharing constraints as it is designed to run in an isolated virtual machine environment. Existing analytical codes can also be integrated with minimal effort.

GEOPHYSICAL INVERSION

A similar approach is used to allow integration with existing third-party and proprietary inversion codes to enable greater reuse of code, transparency, and reproducibility of the research. Geophysical inversions are inherently data- and computationally intensive operations and thus are perfect candidates for execution in the cloud. This provides a high degree of scalability and availability of the services in a cost-efficient manner as well as enabling data security through data and execution environment isolation. The current version of the toolkit incorporated the 1D deterministic sample-by-sample inversion code developed by Geoscience Australia (Brodie, 2015). This code has been released as open-source software and is commonly used in the geophysics community for inverting time-domain electromagnetic data.

CONCLUSIONS

The GPT offers and interactive visualisation environment and infrastructure to link the tool to background processing services. It can support all aspects of geophysical processing for minerals, groundwater, environmental and geo-technical applications. It can be expanded to include ground-based and downhole electromagnetic techniques, as well as other common geophysical surveys such as magnetics, gravity and radiometrics. An interactive workflow offers extensibility interfaces to allow further efficiency improvement of routine tasks undertaken by geophysical data analysts. The GPT implements this approach, which has proven to add value as compared to static data visualisations. It empowers and encourages users to explore data on their own, detect patterns and anomalies in the data and draw informed conclusions.

REFERENCES

1. Brodie, R. C. and Richardson, M., 2015. Open Source Software for 1D Airborne Electromagnetic Inversion. ASEG Extended Abstracts 2015(1), 1-3.
2. Buja A., McDonald J. A., Michalak J. and Stuetzle W., 1991. Interactive data visualization using focusing and linking, Visualization. Visualization '91, Proceedings., IEEE Conference on, San Diego, CA, pp. 156-163, 419. doi: 10.1109/VISUAL.1991.175794.
3. Buja A., Cook D. and Swayne D. F., 1996. Interactive high-dimensional data visualization, Journal of Computational and Graphical Statistics Vol. 5, Iss. 1. doi: 10.1080/10618600.1996.10474696.