GPT: Scalable geophysical survey data analysis and visualisation

Pavel Golodoniuc
CSIRO Mineral Resources
Kensington, WA, Australia
pavel.golodoniuc@csiro.au

Aaron Davis
CSIRO Mineral Resources
Kensington, WA, Australia
aaron.davis@csiro.au

Samuel Bradley
CSIRO Mineral Resources
Kensington, WA, Australia
sam.bradley@csiro.au

Shane Mulè
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.

Key words: geophysics, electromagnetics, data processing, visualisation, exploratory data analysis, human-computer interaction.

INTRODUCTION

Electromagnetic exploration techniques are extensively used for remote detection and measurement of subsurface electrical conductivity structures for a variety of geophysical applications such as mineral exploration and groundwater characterisation. AEM data, which are often originally acquired to answer a specific geophysical question (e.g. green- and brown-field exploration for mineral systems), can be re-used at a later stage to answer different geophysical questions (e.g. groundwater characterisation). This makes AEM data potentially useful beyond their initial purpose for decades into the future.

With surveys ranging from under 1000 km to tens of thousands of kilometres in total length, AEM data are spatially and temporally dense. Sounding stations are often sampled every 0.1-0.2 seconds, with 10-50 measurements taken at each site, resulting in the spacing of measurements along the flight lines of anything between 10-50 metres. This means that the size of a typical AEM survey is in the order of millions of individual stations, with tens of millions of measurements.

Due to intrinsic noise of measurement equipment and noise induced by ground structures (e.g. powerlines, railways, sheds, etc.), AEM data need to be examined for data quality before they can be inverted into the conductivity-depth domain. Data, which are gathered in survey transects or lines, are examined both along the line, in a plan view and for the transient decay of the electromagnetic signal of individual stations before noise artefacts can be removed. Certain noise artefacts can be efficiently detected and removed using numerical methods. Geospatial methods, e.g. proximity to ground structures, can also be employed in conjunction with auxiliary GIS datasets.

The complexity of the data, its size and dimensionality require efficient tools that support interactive visual data analysis and allow easy navigation through the dataset. A suite of numerical algorithms for data quality assurance facilitates this process through efficient visualisations and data quality metrics. The extensible architecture of the Geophysics Processing Toolkit (GPT) allows the application of custom algorithms on-demand through a web-based user interface that seamlessly connects data processing workflow to geophysical inversion codes. Web-based platform for software delivery is chosen consciously to increase accessibility from the widest possible range of devices and operating systems. The toolkit architecture has a small client-side footprint and runs on a standard workstation, delegating all computationally-intensive tasks to the accompanying cloud-based processing unit.

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.

MULTI-VIEW DASHBOARD

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. In our implementation we followed the Visual Information-Seeking Mantra, outlined by Shneiderman (1996): overview first, zoom and filter, then details-on-demand.

Figure 1, at the paper's end, demonstrates the AEM survey visualisation multi-view dashboard that provides an inter-connected 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.

We employ modern rendering techniques, e.g. Scalable Vector Graphics (SVG) and Web Graphics Library (WebGL) along with high-performance data manipulation JavaScript libraries to achieve low latency rendering. Plan view also implements a combination of Douglas-Peucker (Douglas and Peucker, 1973) and Radial Distance algorithms for polyline simplification that significantly reduces the number of points being rendered on the canvas with no visual effect on polyline quality (i.e. shape).

Data-focusing and Cursor Tracking

Dynamic and interactive visualisations encourage even greater engagement with the data when combined with clever user-guided animated transitions. This naturally establishes cognitive links between inspected sections of the data and allows deeper understanding of how data are connected.

Decay profile view offers live cursor tracking functionality for all data channels of a selected flight line in various modes: (a) live cursor tracking for an individual sounding of measured value versus time, and (b) a selection of soundings along the flight line (Figure 2). Selection is automatically reflected in the adjacent view (right) for soundings of decay-value versus time.

COMPLETE 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.

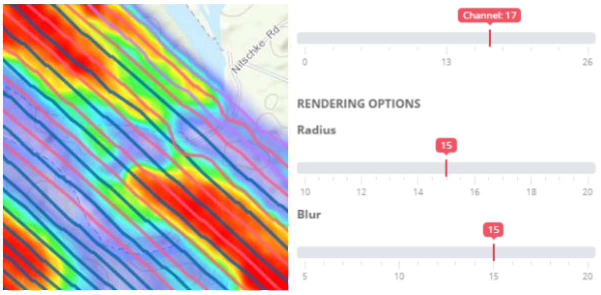

Figure 3. Interactive heatmap plot displaying AEM amplitudes for a selected channel.

We implement various visualisation widgets, such as, a generalised view of AEM survey geophysical properties via an interactive heatmap plot (Figure 3) or highlighting or automatically detected noise. Near real-time performance is achieved with a range of techniques, e.g. dimensionality reduction, use of pre-computed data, caching, etc.

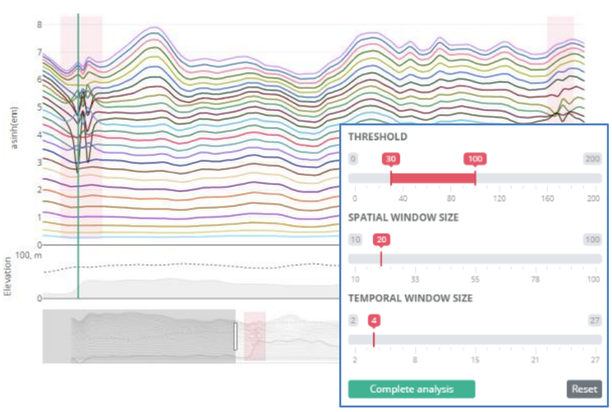

Figure 4. Automatic noise detection parametrization and interactive visualisation of the EM signal.

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 (Figure 4). 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 application programming interfaces (API) 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.

We also envisage support for Jupyter Notebook from within the toolkit, opening endless possibilities for future extensions. This will allow researchers to experiment with new analytical codes using existing survey data live, leveraging the power and scalability of the Cloud platform.

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.

Because of the extensible and open architecture, it is relatively easy to add support for additional inversion codes, processing algorithms and new types of interactive visualisations.

Science-as-a-Service

The toolkit is designed to be integrated with the CSIRO Earth Analytics Industry Innovation Hub (EAIIH) that implements a Science-as-a-Service approach to science delivery for Industry. This inherently provides:

- Data – data discovery and integration are simplified;
- Choice – your data, your algorithm, your parameterisation;
- Scalability – via cloud computing and automation you can explore the whole innovation landscape, to the limits of your time and expense budget;
- Science-as-a-Service – connecting new algorithms into broader business process;
- Lower upfront cost – subscribing to access and ′pay for what you use′ to prove out a new market opportunity.

CONCLUSIONS

The Geophysics Processing Toolkit offers an interactive visualisation environment and an infrastructure in place to link the tool to background processing services, e.g. inversion codes. It can support all aspects of airborne electromagnetic processing for minerals, groundwater, environmental and geo-technical applications. The toolkit 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 intuitive workflow offers extensibility interfaces for future development that improves the efficiency of routine tasks undertaken by geophysical data analysts and interpreters.

Interactive visualisations are at the core of the visual analytics approach to data science and this extends to other data-intensive science domains, including geosciences. The AEM Processing Toolkit implements this approach, which has proven to add value as compared to static data visualisations. Focus, zoom/filter, and drill down methodology empowers and encourages users to explore data on their own, detect patterns and anomalies in the data and draw informed conclusions. Well-crafted visualisation is a great medium for exploratory data analysis, science communication and a mechanism to engage audiences.

Future research will explore 3D visualisation techniques and immersive technologies, e.g. virtual reality (VR) and augmented reality (AR). In conjunction with this, we will conduct a broader study on the human perception of the information presented in these immersive environments.

ACKNOWLEDGEMENTS

The original AEM Processing Toolkit was developed under the Capricorn Distal Footprints project funded by the Science and Industry Endowment Fund (SIEF) as part of The Distal Footprints of Giant Ore Systems: UNCOVER Australia project (RP04-063). It was further supported and advanced via a strategic project development of CSIRO Mineral Resources.

REFERENCES

Brodie, R. C. and Richardson, M., 2015. Open Source Software for 1D Airborne Electromagnetic Inversion. ASEG Extended Abstracts 2015(1), 1-3.

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.

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.

Douglas D. H. and Peucker T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica: The International Journal for Geographic Information and Geovisualization, 10(2), pp.112-122. doi: 10.3138/FM57-6770-U75U-7727.

Shneiderman B., 1996. The Eyes Have It: A task by data type taxonomy for information visualizations. In Proceedings of the IEEE Symposium on Visual Languages, Washington. IEEE Computer Society Press, pp. 336-343

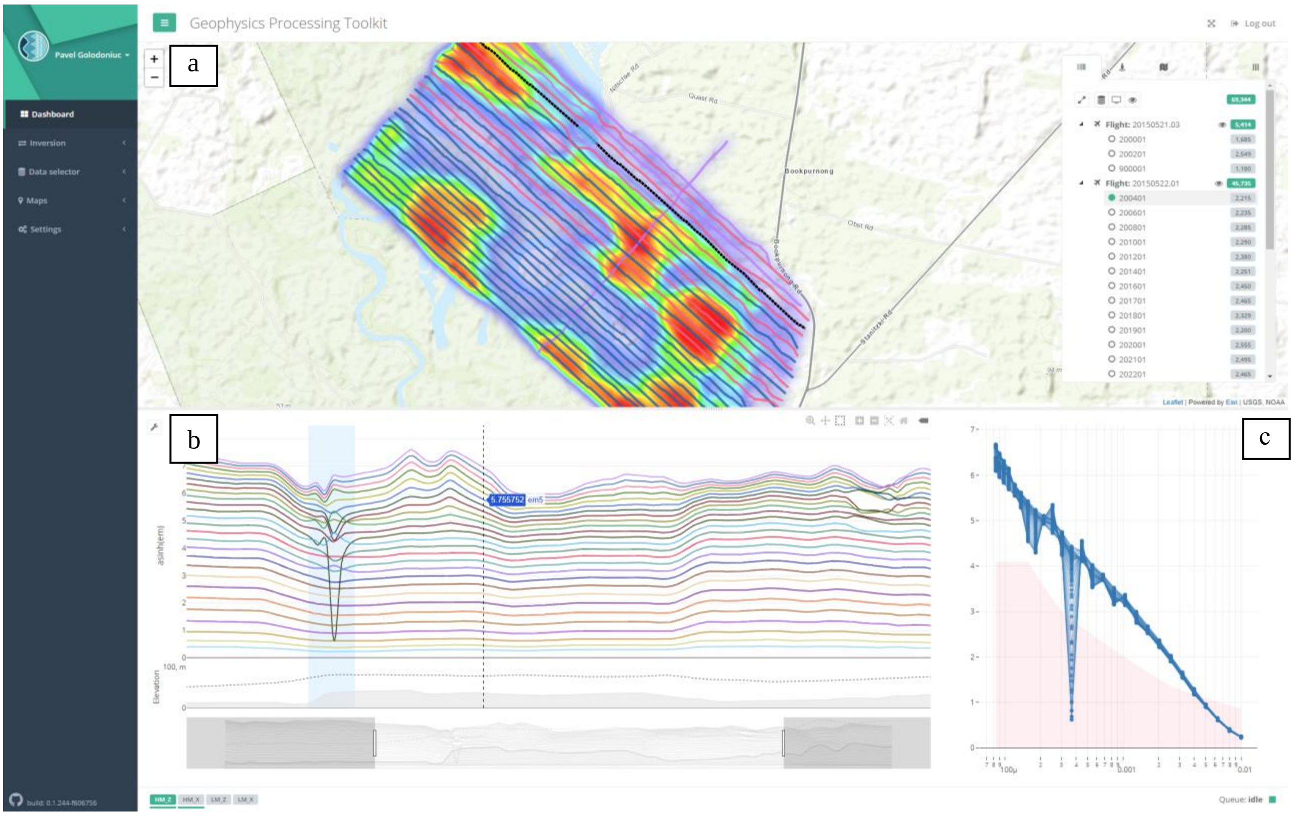

Figure 1. Interactive multi-view dashboard to provide interconnected multi-dimensional visualisation of the AEM survey data: (a) survey plan as a map view with a sidebar outlining survey structure and offering access to additional functionality (base maps, overlays, etc.); (b) fully interactive decay profile showing decays of the transients from individual soundings versus time scale; (c) individual soundings of decay-value versus time for single stations.

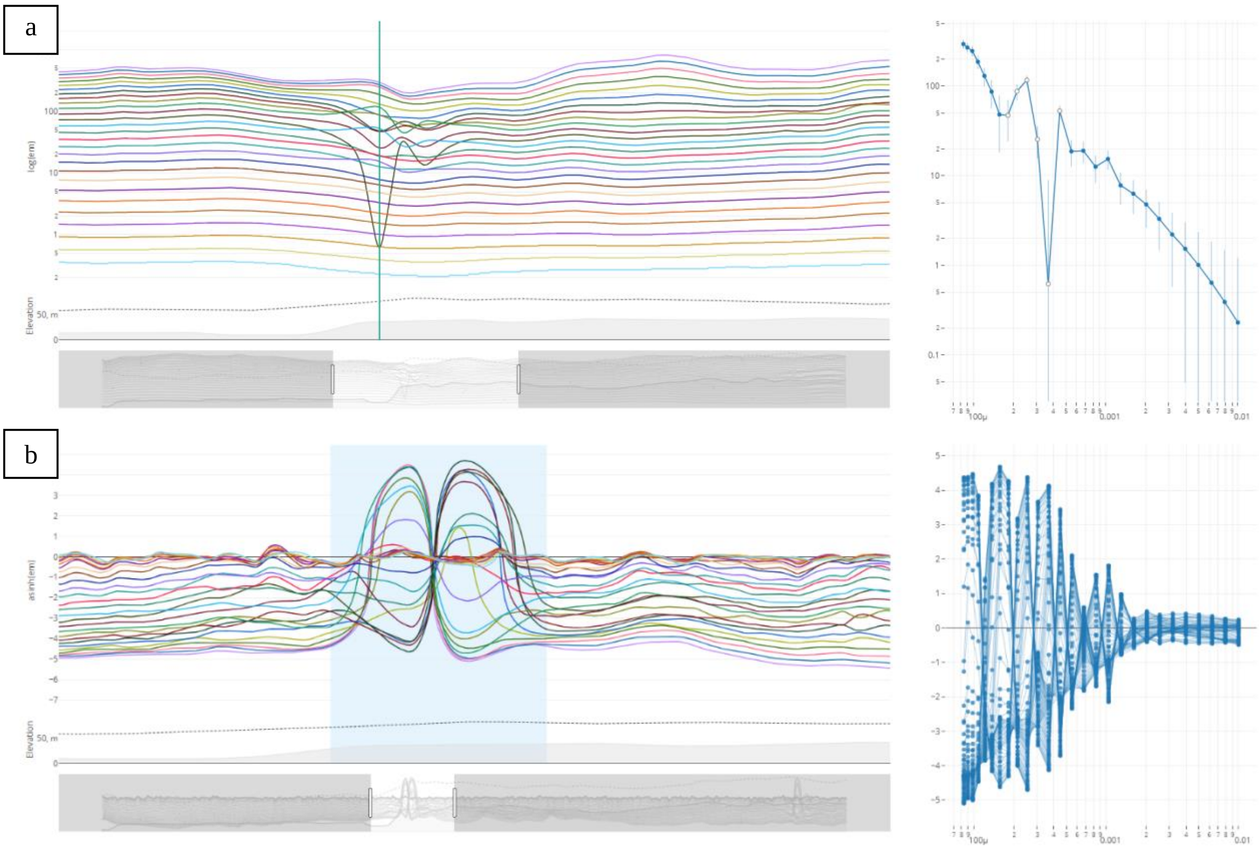

Figure 2. Live cursor tracking over decay profile modes: (a) live cursor tracking – right view implements animated transitions according to the cursor position, (b) a section of sounding (bottom left) and editable decay-value versus time view (bottom right).