Ten years in the wild: The P223 experiment

David Annetts*
CSIRO
26 Dick Perry Av. KENSINGTON 6151
david.annetts@csiro.au

Joe Cucuzza
AMIRA International
Lvl 2, 271 William St, MELBOURNE 3000
joe.cucuzza@amirainternational.com

**presenting author asterisked*

SUMMARY

The use of open-source codes has become pervasive over the past 20 years but such codes are uncommon in minerals exploration. The P223 series of programs researching forward and inverse modelling of electromagnetic data was supported by CSIRO and six AMIRA consortia over 27 years and produced, amongst others, the codes, Airbeo, LeroiAir and Marco. This project concluded in 2008 and, after a two-year embargo, the code base, consisting of computer programs modelling different approximations of the earth for ground and airborne prospecting systems, was released to the public. We discuss reasons why codes have not been more widely adopted, and examine the evolution of some of the codes in research, academia and in industry as a guide to parties who would embark on a similar route.

Key words: open source, electromagnetic modelling, forward, inverse

INTRODUCTION

The use of open-source (OS) computer codes is common in everyday lives. Indeed, one OS operating system currently has the largest installed base of all general purpose operating systems running on platforms from embedded systems to supercomputers. One reason for the ubiquity of such codes is their price. However, other factors such as maintenance and useability are equally important. For example, errors (or bugs) in OS codes are often corrected quicker than in their closed-source counterparts (Anthes, 2016).

It is fair to say that exploration geophysics is dominated by closed-source software. Some inroads have been made in seismic modelling, processing and display by Madagascar, (Fomel et al., 2013), Seismic Unix (Stockwell, 2017) and OpendTect (dGB Earth Sciences, 2017). However, it is difficult to identify counterparts in mining geophysics. One candidate for widespread use in modelling electromagnetic data is the P223 Suite of codes (Raiche et al., 2007). Released to the public in 2010 under a permissive license, this suite has been available to the public for eight years, allowing project sponsors and commercial partners to build on the two-year embargo after the project ended.

As ten years have passed since the P223 project ended, we thought it timely to question whether release of the codes to the public was a success, thereby offering a guide to researchers who would take a similar path. Next, we give an overview of the P223 project.

THE P223 SERIES OF AMIRA PROJECTS

The P223 series of AMIRA projects was carried out at CSIRO between 1980 and 2008. All projects were supported by industry through different AMIRA consortia, each with different numbers of sponsors, and were concerned with the problem of numerically-modelling the electromagnetic response of the earth. Raiche et al. (2007) acknowledge all companies in all consortia in their inversion-focussed review of the final project.

Raiche et al. (2007) summarise each of the codes in some detail. Table 1 provides a brief summary of each of the codes in the suite. Although extensions to multiple-CPU execution are straightforward, codes were provided for single-threaded execution using ANSI-standard F95 compilers and require no external libraries. All codes were designed to be run on high-end desktop computers from a command line and take input from ASCII parameter files (control files). All results are output as ASCII files. A commercial Graphical User Interface (GUI) was developed during the last two P223 projects in order to simplify parameter input and analysis of results (Electromagnetic Imaging Technology, 2017).

Table 1: Brief description of codes provided in the P223 modelling suite. While code to allow Loki to invert data was included in the final project release, further work was required to complete the program. Samaya discretises only a small region of the model and it is recommended that this region have less than 1000 cells. Runtimes are for typical forward models and should be considered indicative.

Ground	Airborne	Inversion	Topography	Runtime (O)	Model
Beowulf	Airbeo	Yes	Flat	Seconds (per station)	1D
Leroi	LeroiAir	Yes	Flat	Minutes	1D + Inductively-thin target
Arjuna	ArjunAir	Yes	Variable	Hours	2.5D
Loki	LokiAir	LokiAir only	Variable	Days	3D
Samays	SamAir	Yes	Variable over small region	Days	Halfspace + 3D region
Marco	MarcoAir	No	Flat	Minutes	1D + Rectilinear prisms

Figure 1 compares code size over the code suite. Calculations of code size omit blank lines and comments. Of the 103 thousand lines of code (kloc) contained by the suite, some 38.4% are contained in AEM codes reflecting the extra complexity of ground-based prospecting systems. Mean ratios of code to comments are 1.6 for the AEM codes and 2.6 for codes modelling ground-based prospecting systems.

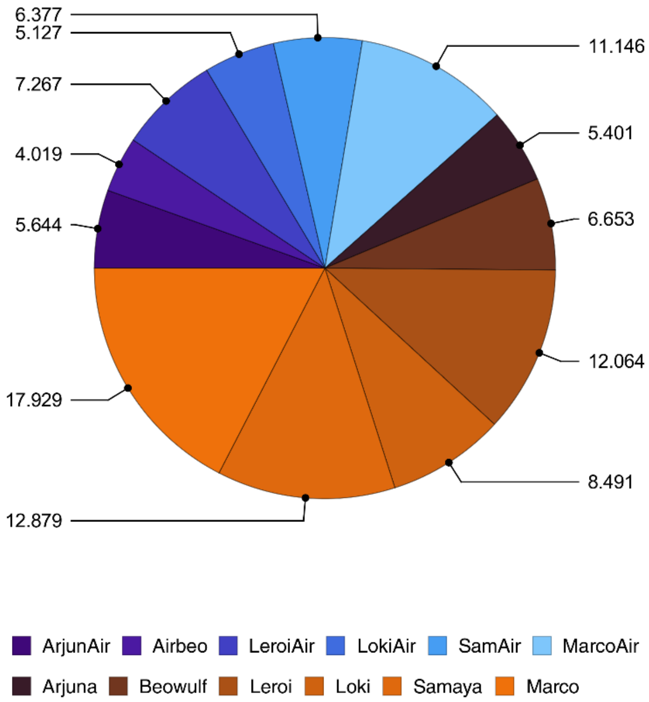

Figure 1: Code size comparison by size. Programs designed to model ground-based prospecting systems comprise 61.6% of the suite. Numbers indicate lines of code (kloc) in a program. These calculations omit blank lines and comment lines.

Code accuracy was of prime concern during code development. Because different codes make different approximations, their comparison is not straightforward. Discretisation requirements further complicate code comparisons, as accurate models can require fine meshes which have large memory requirements, long runtimes or both. Figure 2 compares the response for all airborne electromagnetic (AEM) modelling codes listed in Table 1. Figures 2a and 2b compare the vertical and inline (respectively) component responses for a Tempest AEM system (Lane et al., 2000) while Figure 2e presents the underlying model. The target has a strike length of 1 km. It can be seen that good matches between all codes are achieved at early times. Later than 0.07 ms, ArjunAir responses 2D effects as elevated responses. There is also some evidence from inline responses that the LokiAir model requires refinement. However, such refinement would increase runtimes. SamAir responses are similar in character to other responses but have elevated amplitudes. Should development of the SamAir continue, these differences would need to be addressed. It is also evident that MarcoAir is capable of modelling the 1000:1 conductivity contrast between target and host and that LeroiAir can accurately model the 10 m thick target.

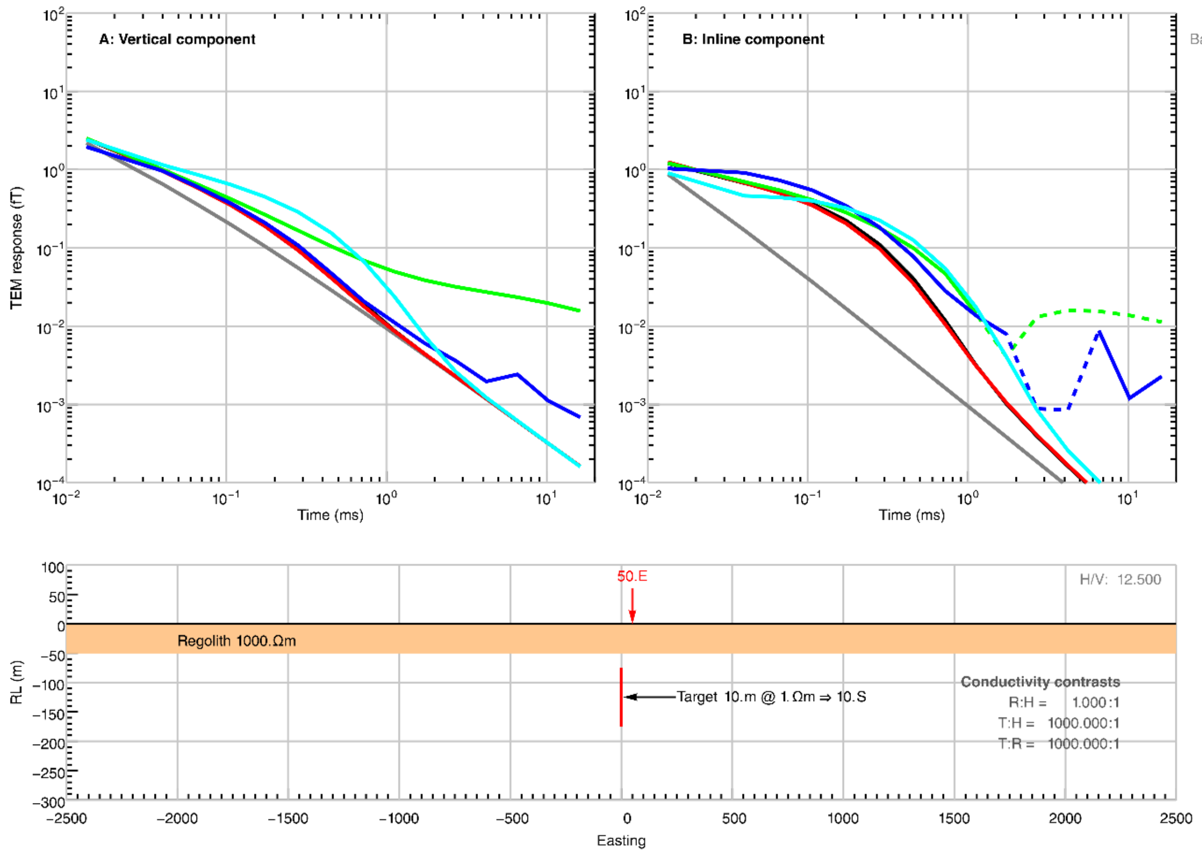

Figure 2: Comparison between P223 AEM codes for a Tempest response of a common model. Panel A compares vertical-component data while panel B compares inline-component data. Codes are in good agreement at early times. After 0.07 ms, 2.5D effects are evident in ArjunAir responses. LokiAir responses suffer at late time from too coarse a mesh. It is also evident that in some circumstances, MarcoAir can accurately model high-contrast targets and that LeroiAir is not necessarily limited to inductively-thin targets.

POST-PROJECT CODE USE

Post-project code use is difficult to quantify. There is no guarantee that downloaded code is used, and no guarantee its use is reported. For example, there is no obligation by universities employing the codes for teaching that they report that use. Nevertheless, Sourceforge (2010) reports 357 downloads of the code suite.

Another measure of post-project code use is can be made by a literature search for codes released as part of the suite. This measure is imprecise because it relies on work being published; conference presentations that do not require extended abstracts which do not appear in such a search. By this metric, some 87 papers cite codes contained by the P223 modelling suite. Unfortunately, not all these references can be summarised in the space of an extended abstract. Accordingly, we briefly describe two applications that have been built from the P223 code release.

Pirttijärvi and Salmirinne (2015) describe modifications to ArjunAir. These modifications take the form of a graphical user interface and a different inversion philosophy to the original. ArjunGUI modifies the Jacobian and replaces the original SVD-based inversion (Jupp and Vozoff, 1975) with two alternatives. One of these is also SVD-based but employs adaptive damping while the other implements constrained Occam (Constable et al., 1987) inversion. Other modifications made by Pirttijärvi and Salmirinne (2015) include removal of scratch memory and the use of OpenMP parallelisation. Like the P223 code base, ArjunGUI is currently open source software.

In a series of articles, Hauser at al (2015; 2016; 2017) describe another evolution of two codes from the P223 code base. Recognising that a key factor in products derived from airborne AEM data is survey line spacing, Hauser et al (2015) pixelate a tenement-sized area of a regional survey, then employ a spatially-constrained Bayesian Parametric Bootstrap (BPB) to invert multiple data sets for geologically-relevant interfaces and basement conductors using layered-earth code from AirBeo and discrete conductor code from LeroiAir. Hauser et al (2017) describe the use of these interfaces in groundwater modelling. Hauser et al (2016) extended their stochastic approach to the search for basement conductors.

DISCUSSION

Although a list of 87 papers citing codes contained in the P223 code suite is encouraging, it is not necessarily an indication of the success of the suite as an OS-code. Different authors have used different measures to determine the success of such projects. Ghapanchi et al. (2011) employ five criteria to determine the success of OS projects. Their criteria relate mainly to large projects which comprise a number of developers rather than to single code releases. The only criterion of Ghapanchi et al. (2011) that is directly applicable is ‘user interest’, and this is reflected in papers citing the codes.

Another perspective on the success of OS projects is provided by Schweik (2013) who studied nearly 174000 OS projects, in the process, surveying around 1400 developers. Schweik (2013) posed the question “What factors lead to some OS commons to success and others to abandonment”. Of the 174000 projects analysed, only 15% achieved success in growth which was defined as three meaningful releases deemed useful for a few users. Some 35% were considered abandoned at initiation where no first release was produced. By Schweik’s (2013) criteria, we consider the P223 code base indeterminate in growth since although no more than three releases have been produced, there has been some development activity.

Schweik (2013) suggests that factors such as the creation of a ‘virtuous circle’ between users and developers, scheduled tasks such as code enhancements and bug fixes, competition, financing and administration through code revision systems can improve conditions for developers and users. Subramaniam et al. (2009) have suggested that licensing is an important factor in success; restrictive licenses had a negative effect on project success when projects were targeted at software developers. Project growth is a natural consequence of these improved conditions.

Common components of successful OS projects (Subramaniam et al., 2009; Schweik, 2013) are developer activity, user interest and program activity. Clearly, there is user interest in the P223 Suite. Activity by users, in the form of citations of the code is steady. Perhaps more encouragingly, components of the suite have been used to develop applications (Silic et al., 2015; Pirttijärvi and Salmirinne, 2015; Hauser et al., 2016; Kiyan et al., 2017).

CONCLUSIONS

Since the project ended in 2008, the P223 suite of electromagnetic modelling codes has seen low-level interest. This interest has mainly been in the form of code use. However, there has been some development as new applications are built from P223 codes. For both reasons, we suggest that release of the P223 modelling suite has been successful. Partly because it was never released as such, the P223 code base has yet to transition from code release to OS project. We suggest that such a transition is possible, provided activity by developers and users is encouraged towards this end.

ACKNOWLEDGMENTS

This abstract would not be possible without the support of industry sponsors of the series of AMIRA P223 projects. We further thank Juerg Hauser, Brett Harris, Kim Frankcombe, Ron Palmer and Volker Rath for discussions around their use of codes in the P223 code suite.

REFERENCES

Anthes, G., 2016, Open Source Software No Longer Optional: Communications of the ACM, 59(8), 15–17.

Constable, S.C., Parker, R.L. and Constable, C.G., 1987, Occam’s inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, 52(3), 289–300.

dGB Earth Sciences, 2017, OpendTect: OpendTect. Available at: http://www.opendtect.org/ [Accessed August 14, 2017].
Electromagnetic Imaging Technology, 2017, Maxwell: Available at: https://www.electromag.com.au/maxwell.php [Accessed August 25, 2017].

Fomel, S., Sava, P., Vlad, I., Liu, Y. and Bashkadin, V., 2013, Madagascar: open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, 1(1), E8.

Ghapanchi, A.H., Aurum, A. and Low, G., 2011, A taxonomy for measuring the success of open source software projects: First Monday. Available at: http://firstmonday.org/ojs/index.php/fm/article/view/3558/3033 [Accessed August 7, 2017].

Hauser, J., Gunning, J. and Annetts, D., 2016, Probabilistic inversion of airborne electromagnetic data for basement conductors: Geophysics, 81(5), E389–E400.

Hauser, J., Gunning, J. and Annetts, D., 2015, Probabilistic inversion of airborne electromagnetic data under spatial constraints: Geophysics, 80(2), E135–E146.

Hauser, J., Wellmann, J.F. and Trefry, M., 2017, Water table uncertainties due to uncertainties in structure and properties of an unconfined aquifer: Groundwater, DOI: 10.1111/gwat.12577.

Jupp, D.L.B. and Vozoff, K., 1975, Stable Iterative Methods for the inversion of Geophysical Data: Geophysical Journal of the Royal Astronomical Society, 42, 957–976.

Kiyan, D., Rath, V. and Delhaye, R., 2017, An inversion toolbox for frequency- and time-domain airborne electromagnetic data from surveys in Ireland, in Second European Airborne Electromagnetics Conference: EAGE.

Lane, R., Green, A., Golding, C., Owers, M., Pik, P., Plunkett, C., Sattel, D. and Thorn, B., 2000, An example of 3D conductivity mapping using the TEMPEST airborne electromagnetic system: Exploration Geophysics, 31(2), 162–172.

Pirttijärvi, M. and Salmirinne, H., 2015, ArjunGUI: 2D modelling and inversion software for airborne time-domain EM data, in Novel Technologies for Greenfield Exploration: Geological Survey of Finland, p. 199.

Raiche, A., Wilson, G. and Sugeng, F., 2007, Practical 3D inversion -- The P223F Software Suite, in 19th Geophysical Conference and Exhibition. ASEG, pp. 68–69.

Schweik, C.M., 2013, Sustainability in Open Source Software Commons: Lessons Learned from an Empirical Study of SourceForge Projects: Technology Innovation Management Review, 3(1), 13–19.

Silic, J., Paterson, R., FitzGerald, D.J. and Archer, T., 2015, Comparing 1D and 2.5D AEM Inversions Using a New Adaptive Inversion Solver in a 3D Geological Mapping Environment, in Near Surface Geoscience.

Sourceforge, 2010, P223Suite: P223Suite. Available at: https://sourceforge.net/projects/p223suite/?source=directory [Accessed August 25, 2017].

Stockwell, J., 2017, CMP/SU: Seismic Unix: CWP. Available at: http://www.cwp.mines.edu/cwpcodes/ [Accessed August 14, 2017].

Subramaniam, C., Sen, R. and Nelson, M.L., 2009, Determinants of open source software project success: A longitudinal study: Decision Support Systems, 46(2), 576–585.