COFI: A COmmon Framework for Inversion

All our knowledge of the Earth's interior structure, internal processes, dynamics and geological history, as well as its physical and chemical evolution, has been derived from observations made at or near the Earth's surface combined with laboratory studies. Observations are always indirect and some form of inference process must be used to see what's hidden from view beneath the surface. So too in the laboratory, where we measure what is available today, but want to know what happened in the past. We cannot directly access the deep interior, nor observe Earth's past structure, and always have only partial information. The geoscientist is therefore reliant on learning from indirect measurements. This creates what is known as an *inverse or inference problem*, where we learn by combining knowledge of physics and chemistry together with observations, allowing us to decipher the information contained in complex signals from the planet.

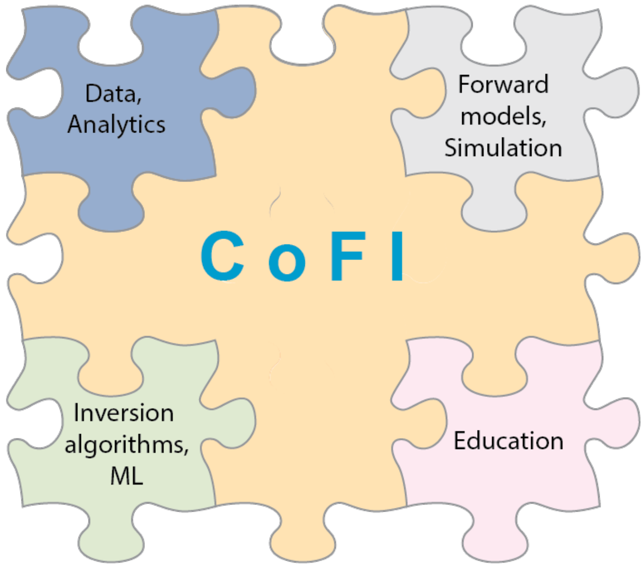
Learning from indirect observations is crucial to answering fundamental questions about how the Earth works and what its future will be, as well as addressing present-day issues of crucial societal and industrial importance. For example, analysis of indirect geophysical data underpins all hydro-carbon and mineral exploration, as well as the location of earthquakes, tectonic plate reconstructions and present day motions, while geological and geochemical information indirectly constrains Earth's complex interior evolution. Progress in all areas depends on the quality and sophistication of inferences that can robustly made with complex datasets. Inverse problems also appear in many other areas of the physical, life and mathematical sciences, with essentially similar underlying mathematical structure. Furthermore, artificial intelligence and machine learning algorithms – whose transformative potential has been widely recognized – have at their heart the solution of inference problems which are closely related to those arising across the Earth Sciences.

The problem

With inference problems lying at the heart of a diversity of scientific applications it is no surprise that methods, algorithms and tools arising from one field find application across scientific boundaries. Many scientists are aware that such cross fertilization can be the key to solving their own challenging research questions, and can create new capability where none previously existed. Given the recognized importance of such 'cross-talk' there are surprisingly few mechanisms to ease the path. Take up of new data analysis methods often happens in a rather ad hoc and piecemeal manner. The growing appreciation of machine learning (ML) across the sciences is a pertinent example: geoscientists often lack the background and training to properly appreciate the underlying assumptions inherent in ML methods, their potential and their limitations. More generally, there are significant barriers to the accessibility of advanced inference tools, in terms of software, awareness and education, which inhibits their use across different sub-disciplines.

A way forward

What is lacking is a common language, or framework, that can allow researchers in varied disciplines to communicate. Those who wish to perform inference confront a barrier in defining their problems with sufficient rigor to connect with those who develop solutions. A common language would allow users to describe the characteristics of their problem in a way that makes it accessible to researchers without domain-specific knowledge, and to straightforwardly identify and adopt potentially-applicable methods of solution. Conversely, testing and benchmarking of new algorithms and techniques could be greatly simplified by easy access to a suite of problems of different characteristics and scales. In principal each component of data, uncertainty, mathematical model, computational simulation, and inference method could be expressed within the framework of a common interface, like pieces of a jigsaw that fit together. In this way the power of the next generation inversion, machine learning and statistical inference algorithms could be made accessible through purpose-built computational infrastructure connecting users and developers. COFI could be that open standard for inference. Educational resources could then be developed across an increasingly wide set of disciplines connecting data questions to sophisticated solvers without the need for reinvention or repeated local tailoring of solutions. Researchers could be connected into a community network of practitioners who can help span discipline barriers and solve problems through a common protocol.



A common (open source) framework for inversion would make it possible to seamlessly connect data, forward models, inversion tools, for research and education.

What would it look like?

COFI would most likely consist of an open standard for specifying the components of each type of inference problem: data type, parameterization, forward model making predictions from a solution, inverse solver estimating the solution from the data together with uncertainty, and so on. Building on this, it would define an API (application programming interface) allowing disparate software components to connect to one another in a standardized way. By implementing this API within their software, end users would be able to seamlessly interface with any other COFI-standard forward model or inverse solver. We envisage a framework capable of facilitating the solution of inverse problems from the simplest to the most complex; from textbook solutions to cutting edge algorithms; from the laptop to HPC. As different disciplines build interfaces to COFI then the number of applications, solvers, datasets and hence utility would grow facilitating a worldwide community resource for the solution of all classes of inverse problem.

Why start in Australia?

Many individuals and organizations with the required expertise in this area already exist within the Australian region based in university, government research organizations or industry. Leading researchers in inversion and geophysical imaging exist across Australian institutions, while a larger number of researchers are currently in need of advanced data inference, inversion and increasingly machine learning algorithms. Australia also has world leading funding mechanisms for software infrastructure through NCRIS (AuScope), and via partnerships between universities, industry and institutions such as the CSIRO and Geoscience Australia.

A COFI club

We envisage building a community of interested parties, initially with a geoscience focus, with expertise spanning data collection to algorithm development, software and infrastructure to advise on the shape that COFI may take, and help make it happen, raise supporting funds, devote time and influence its direction. Now is an opportune time to start that conversation.