PH.D. THESIS PROPOSAL:

A Framework for Geophysical Inversions

Rowan Cockett

May 04, 2016

Supervisor:
Eldad HABER

External Examiner:
Craig HART

Committee Members:
Doug OLDENBURG
Roger BECKIE

Examination Chair:
Douw STEYN

Department of Earth, Ocean and Atmospheric Sciences

The University of British Columbia

Contents

1 Introduction

1.1 Motivation

The goal of the applied geosciences is to make predictions and decisions about the subsurface. The accuracy of these predictions can have far reaching economic and environmental implications (e.g. contamination delineation, resource exploration, reservoir optimization). There are many disciplines and skills that are involved in providing predictions and increasingly these disciplines must collaborate and integrate their domain specific knowledge. In a managed aquifer recharge project, for example, the goal is to infiltrate water into the subsurface for storage and subsequent recovery. Throughout the lifetime of the project monitoring and management of the infiltration site is necessary (e.g. Racz et al. (2011); Daily et al. (1992); Park (1998)). Such projects require input from geology, hydrology and geophysics to map the hydrostratigraphy, to collect and interpret time-lapse geophysical measurements, and to integrate all results to make predictions and decisions about fluid movement at the site. The integration of the geosciences is far from trivial as each discipline has differing descriptive terminology as well as software tools that are domain specific with limited interoperability. Each of these domains involves their own physics problems and versions of discretization, regularization, optimization, computer science, empirical relations, and geology. In the geosciences, progress in integrating disciplines is hindered by an unformalized ontology and a history of proprietary software tools that have limited accessibility and poor interoperability. Geoscience integration is currently the target of major funding initiatives (e.g. EarthCube - 11 year NSF project, 35M in 2015; CIMIC Footprints Project - NSERC Project, 24 Universities, 30 Industry, 13M). Many of the current efforts are focused on software infrastructure, formally describing geoscientific data (using ontologies), and formally describing methods of integrating disciplines. Integration of disciplines and formally promoting quantitative communication throughout geophysical simulations and parameter estimation is crucial to integration; this is the main focus of my PhD.

1.2 Geophysical inversion framework

Geoscience inversions are the mathematical process of creating subsurface models to fit measured data, generally guided by physics. The language, workflow, and resulting software implementations of geoscience inversions varies across disciplines. These inconsistencies are among the large barriers to sustained cross-disciplinary integrations. To address this barrier to geoscience integration, I strive to formalize geoscience inversions. This process will take the form of deriving, from the existing body of literature, a consistent terminology, a conceptual framework, and a software organization, which supports reproducible inversion workflows. By formalize, I do not mean mathematically, rather taking practices of ontology and framework development in biology and other more mature interdisciplinary fields and applying them to geophysical inversions. In a Nature genetics review by Bard and Rhee (2004) the authors conclude: *"An ontology makes explicit knowledge that is usually diffusely embedded in notebooks, textbooks and journals or just held in academic memories, and therefore represents a formalization of the current state of a field."*. Adapting these methodologies to geophysical inversions inherently requires that a diverse suite of methods and applications be considered across geophysics, hydrogeology and geology. The geoscience problems that I am using to illuminate this work are: direct current resistivity (Cockett and Pidlisecky, 2014; Pidlisecky et al., 2013; Steklova et al., 2012), frequency/time domain electromagnetics (Kang et al., 2014;

Heagy et al., 2015c; Fohring et al., 2012; Kang et al., 2015a), vadose zone fluid flow (Cockett & Haber, in prep), and parameterized geologic modeling (Cockett et al., 2016; Cockett, 2013, 2012). The way in which I am working to formalize geoscience inversions is through the development of a rigorously tested framework that supports and enables these works. Each of these problems has its own unique constraints: each has its own physics and physical properties, empirical relations may be considered, requirements on the spatial and temporal discretizations, large matrix systems to solve with variable structure and possibly with many sources, and unique parameterizations that are context specific for a geoscience decision.

The initial version of the framework was published in Computers and Geoscience (Cockett et al., 2015c), a brief illustration of the major components and layout of the framework is seen in Figure 1, for a more in depth discussion, please refer to the paper, the online documentation (`http://docs.simpeg.xyz`) or the primary open-source numerical implementation (`https://github.com/simpeg/simpeg`). The framework has since been extended, refined and used for research problems in: induced polarization (Kang and Oldenburg, 2016), electromagnetics (Heagy, Cockett, et al; in prep), magnetotellurics (Rosenkjaer et al. in prep, Rosenkjaer et al. (2015b)), gravity, magnetics, and seismic, as well as a number of other conference proceedings (Cockett et al., 2015a; Rosenkjaer et al., 2015a). The framework and software has also seen use in industry applications, as well as university courses and training (Heagy et al., 2015a,b), in reproducible textbooks (http://em.geosci.xyz, http://gpg.geosci.xyz), has seen multiple awards (UBC Library Award, NSERC Doctoral Prize for Interdisciplinary Research) and will be the subject of an upcoming conference (BIRS - Geophysical Simulation and Inversion, August 2016).

I plan to address and further develop this framework through two avenues:

1. vadose zone inverse problems using Richards equation with careful attention to the computational scalability of the parameter estimation problem;

2. integration of simple process based geologic parameterizations with the geophysics framework to support geologic interpretation of geophysical data.

Additionally, in collaboration with Heagy, Kang, Rosenkjaer and Oldenburg I will extend and refine the framework to support electromagnetic inversions in a variety of formulations and applications of steel casing and geothermal exploration. The work on the electromagnetics framework is to be submitted by June 2016 to Geophysics. These research areas may seem disparate, but collectively they are united by a common theme of addressing the complexities of bringing together the disciplines of geophysics, hydrology, geology, and inverse theory in a computationally tractable manner that is accessible and reproducible by other researchers. In this proposal I will provide a brief background on geophysical inverse problems and vadose zone flow, highlighting the need for computational feasibility and a framework to support simulation and inversion across multiple disciplines.

Throughout this work, my colleagues and I have tried to summarize and/or reproduce a large quantity of the geophysical inversion literature. This has been difficult due to the diverse terminology describing extremely similar methodologies and a lack of numerical instructions (often intentionally withheld). This proprietary paradigm does not enable the diverse collaborations necessary for geoscience integration. I have looked to and learned from other fields (e.g. biology, astrophysics, machine learning Astropy Collaboration et al. (2013); Pedregosa et al. (2011)) on how to organize a community around open, accessible, actionable ideas. Adapting these learnings to geoscience, I strive to complete all of my research such that it is immediately reproducible and openly available (Cockett et al., 2015b; Wilson et al., 2014; Holscher et al., 2010; Kalderimis and
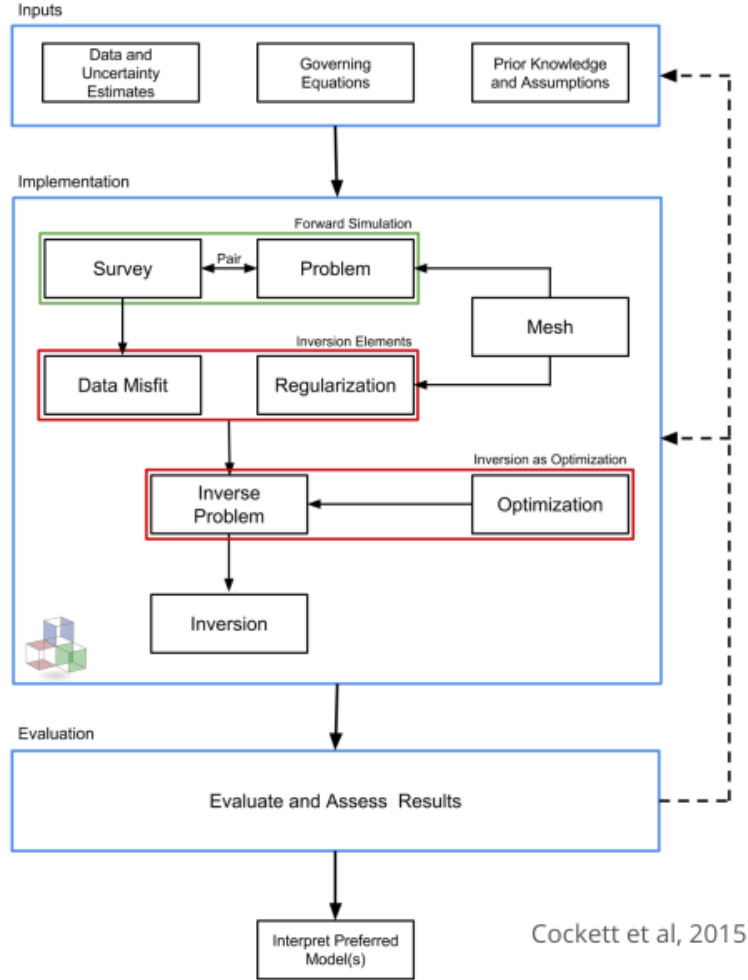
Figure 1: Illustration of the SIMPEG framework showing the major components and their inter-connectivity. Adapted from Cockett et al. (2015c) (CC-BY-4.0).

Meyer, 2011; Merwin et al., 2015). Much of my work requires software, which is 'implementation' and 'engineering' by its nature. However, the software (although extremely useful) is not the aim of my research: software is the means by which I test, extend, and generalize my ideas in a rigorous, scientific way. The aim of my research is to identify, explore, and formalize a framework for simulation and parameter estimation in geophysics.

1.3   Application to vadose zone fluid flow

The development of a geophysical framework requires considering a number of disciplines and geophysical problems to ensure generality as well as extensibility. I am working with collaborators in many of these geophysical methods (electromagnetics, direct current resistivity, magnetotellurics, magnetics, gravity) and am ensuring that the framework that I am developing supports these applications. However, the goal is also to have the framework work outside of geophysics and most notably in hydrogeology, as such, I have chosen vadose zone fluid flow as a model problem. Addi-

4

tionally, there are a few numerical techniques which I have identified that would be a significant contribution in the context of hydrogeophysical inversions. Fluid flow in the vadose zone is governed by Richards equation and parameterized by hydraulic conductivity, which is a nonlinear function of pressure head. Investigations in the vadose zone typically require identification of distributed hydraulic properties. This is increasingly being done through a data-driven approach, using changes in saturation or pressure head to infer hydraulic parameters. Hydrogeophysics allows many more proxy measurements, such as direct current resistivity data, to be taken to help characterize these sites spatially, as well as through time. As such, the number of distributed hydraulic parameters to be estimated in a Richards equation inversion will continue to grow. There has been much research into the scalability of the inverse problem in geophysical applications (e.g. electromagnetics) that allow the calculation of an optimization step in the inversion *without* explicitly calculating or storing the sensitivity matrix (cf. Haber et al. (2000)). This is extremely important in large problems as the computer memory available to store this large, dense matrix can often be a limitation. Previous work use either numerical differentiation, automatic differentiation, or finite difference in order to compute the sensitivity matrix (e.g. PEST) (Finsterle and Zhang, 2011; Bitterlich and Knabner, 2002). This is computationally slower and can generate inaccuracies in the sensitivity computation and tarry convergence of the optimization algorithm. With regard to implicit sensitivity calculations, there is an opportunity to apply some of the learnings from the geophysical inversion literature to this hydrogeologic problem. Note that the sensitivity calculation is necessary in *any* gradient based technique.

The application of the implicit sensitivity calculation to Richards equation, however, is not straightforward. Hydraulic conductivity is the function we are inverting for - it is an empirical function that depends on pressure head - pressure head is the field that is calculated using Richards equation. This nonlinear coupling requires iterative optimization methods in the forward simulation between each time step (e.g. Picard or Newton). This makes the implicit calculation of the effect of the sensitivity on a vector rather involved and intricate. Furthermore, the nonlinear relationship of hydraulic conductivity is empirically determined and depending on the relation used could involve the estimation of up to nine spatially-distributed parameters from a finite dataset. The implicit use of the sensitivity matrix should have the ability to calculate the sensitivity to any of these model parameters, however, any inversion algorithm would be ill-advised to try to estimate all distributed parameters at once. One goal of the proposed work is to tackle the sensitivity calculation *implicitly*. This would further allow for exploration into different inversion methodologies and parameterizations of the empirical relationships. By not storing the sensitivity, and instead computing its effect on a vector, the size of the problem that we can invert will become much larger. This will allow large 3D hydraulic parameter inversions using Richards equation to be run on modest computational resources. Below is a brief mathematical sketch of the implicit sensitivity calculation. The full derivation as well as the numerical implementation are to be submitted in June 2016 to Inverse Problems.

2 Geophysical inverse problems

2.1 Background and motivation

Geophysical surveys can be used to obtain information about the subsurface as the responses that are measured depend on the physical properties and contrasts in the earth. Inversions provide a

mathematical framework for constructing subsurface models consistent with the data collected by these surveys. The data collected are finite in number while the physical property distribution of the earth is continuous. Thus, inverting for a physical property model from geophysical data is an ill-posed problem, meaning that no unique solution explains the data. Furthermore, the data may be contaminated with noise. Therefore, the goal of a deterministic inversion is not only to find a model consistent with the data, but must be to find the 'best' model that is consistent with the data. The definition of 'best' requires the incorporation of assumptions and *a priori* information, often in the form of an understanding of the particular geologic setting or structures (Constable et al., 1987; Oldenburg and Li, 2005a; Lelièvre et al., 2009). Solving the inverse problem involves many moving pieces that must work together, including physical simulations, optimization, linear algebra, and incorporation of geology. Deterministic geophysical inversions have been extensively studied, and many components and methodologies have become standard practice. With increases in computational power and instrumentation quality, there is a greater drive to extract more information from the geophysical data. Additionally, geophysical surveys are being applied in progressively more challenging environments. As a result, the geosciences are moving towards the integration of geological, geophysical, and hydrological information to better characterize the subsurface (e.g. (Haber and Oldenburg, 1997; Doetsch et al., 2010; Gao et al., 2012)). This is a scientifically and practically challenging task (Li and Oldenburg, 2000; Lelièvre et al., 2009). These challenges, compounded with inconsistencies between different data sets, often makes the integration and implementation complicated and/or non-reproducible. The development of new methodologies to address these challenges will build upon, as well as augment, standard practices; this presupposes that researchers have access to consistent, well-tested tools that can be extended, adapted and combined.

There are many proprietary codes available that focus on efficient algorithms and are optimized for a specific geophysical application (e.g. (Kelbert et al., 2014; Li and Key, 2007; Li and Oldenburg, 1996, 1998)). These packages are effective for their intended application, for example, a domain specific large-scale geophysical inversion or a tailored industry workflow. However, many of these packages are 'black-box' algorithms, that is, they cannot easily be interrogated or extended. As researchers, we require the ability to interrogate and extend ideas; this must be afforded by the tools that we use. Accessibility and extensibility with regard to the geoscience application are the primary motivators for my work. Other disciplines have approached the development of these tools through open source initiatives using interpreted languages, for example, Astropy in astronomy (Astropy Collaboration et al., 2013) and SciPy in numerical computing (Jones et al., 2001). Interpreted languages facilitate interactive development using scripting, visualization, testing, and interoperability with code in compiled languages and existing libraries. Furthermore, many open source initiatives have led to communities with hundreds of researchers contributing and collaborating using social coding platforms, such as GitHub (https://github.com). There are also initiatives in the geophysical forward and inverse modeling community targeting specific geophysical applications (cf. (Hansen et al., 2013; Hewett et al., 2013; Uieda et al., 2014; Kelbert et al., 2014; Harbaugh, 2005)). I am interested in creating a community around geophysical simulations and gradient based inversions that is contributing to a consistent body of knowledge. To create a foundation on which to build a community, we require a comprehensive framework that is applicable across domains and upon which researchers can readily develop their own tools and methodologies. To support these goals, this framework must be modular and its implementation must be easily extensible by domain specific researchers.

2.2 Take home points

Sustained, reproducible integration requires that methodologies be accessible, consistent, numerically documented, interoperable and extensible. To do this, research is required to:

- Identify the pieces and interfaces

- Abstract commonalities to a consistent, supporting subset

- Capture heuristics in a reproducible manner

- Numerically test, capture and document

The output of this will be a conceptual framework that is numerically tested and demonstrates the capability to support existing and future research directions, and will ideally catalyze and accelerate interdisciplinary collaborations.

3 Vadose Zone Flow

3.1 Motivation

Studying the processes that occur in the vadose zone, the region between the earth's surface and the fully saturated zone, is of critical importance for the understanding of our groundwater resources. The majority of groundwater recharge is derived through water that percolates through the vadose zone. As such, much attention has been given to monitoring and describing processes that reside in this region of the earth. Flow in the vadose zone is regulated by hydraulic conductivity, which is a non-linear function of saturation. This non-linear function is defined empirically using, for example, the Brooks-Corey model (Brooks and Corey, 1964) or the van Genuchten-Mualem model (Mualem, 1976; van Genuchten, 1980). Typically, the hydraulic conductivity is heterogeneous and can have large dynamical range. The spatial estimation of the hydraulic conductivity function is an important step in any site characterization; however, this estimation is difficult and simplifications are consistently used to avert these conceptual and computational difficulties. These simplifications typically parameterize the conductivity and assume that it is a simple function in space (e.g. Liang and Uchida (2014)). Such assumptions are often sufficient to fit observed (or conceptual) measurements because of the lack of constraining data, especially in two- and three-dimensions as well as in time. As more data becomes available through spatially extensive surveys, and time-lapse proxy measurements (e.g. direct current resistivity surveys, distributed temperature sensing) the possibility to extract more information about subsurface hydrogeologic parameters becomes a possibility. Fluid flow in the vadose zone is governed by Richards equation and parameterized by hydraulic conductivity, which is a non-linear function of pressure head. Recent advancements have been made to simulate Richards equation in a computationally scalable manner (Orgogozo et al., 2014). However, the inverse problem is non-trivial, especially in three-dimensions, and must be considered using modern numerical techniques that allow for spatial estimation of hydraulic parameters.

3.2 Background

Flow in the vadose zone has many complications as the parameters that control the flow are dependent on the saturation of the media, leading to a non-linear problem. The groundwater flow

equation has a diffusion term, as well as an advection term that is related to gravity and only acts in the z-direction. There are two different forms of Richards equation that differ on how they deal with the non-linearity in the time-stepping term. Here we use the most fundamental form, referred to as the 'mixed'-form of Richards equation (Celia et al., 1990):

$$\frac{\partial \theta(\psi)}{\partial t} - \nabla \cdot k(\psi)\nabla\psi - \frac{\partial k(\psi)}{\partial z} = 0 \quad \psi \in \Omega \tag{1}$$

where θ is water content, and ψ is pressure head. This formulation of Richards equation is called the 'mixed'-form because the equation is parameterized in ψ but the time-stepping is in terms of θ. The hydraulic conductivity, $k(\psi)$, is heterogeneous and potentially anisotropic function that is assumed to be given when solving the forward problem. In the following equations we assume that k is isotropic, but the extension to anisotropy is straightforward and is completed numerically. The equation is solved in a domain Ω equipped with boundary conditions on $\partial\Omega$ and initial conditions that are problem dependent.

An important aspect of unsaturated flow is noticing that both water content, θ, and hydraulic conductivity, k are functions of pressure-head, ψ. There are many empirical relations used to relate these parameters, including the Brooks-Corey model (Brooks and Corey, 1964) and the van Genuchten-Mualem model (Mualem, 1976; van Genuchten, 1980). The van Genuchten model is written as:

$$\theta(\psi) = \begin{cases} \theta_r + \dfrac{\theta_s - \theta_r}{(1 + |\alpha\psi|^n)^m} & \psi < 0 \\ \theta_s & \psi \geq 0 \end{cases} \tag{2a}$$

$$k(\psi) = \begin{cases} K_s\theta_e(\psi)^l(1 - (1 - \theta_e(\psi)^{-m})^m)^2 & \psi < 0 \\ K_s & \psi \geq 0 \end{cases} \tag{2b}$$

where

$$\theta_e(\psi) = \frac{\theta(\psi) - \theta_r}{\theta_s - \theta_r}, \qquad m = 1 - \frac{1}{n}, \qquad n > 1 \tag{3}$$

Here θ_r and θ_s are the residual and saturated moisture contents; K_s is the saturated hydraulic conductivity; α and n are fitting parameters; and $\theta_e(\psi) \in [0, 1]$ is the effective saturation. The pore connectivity parameter, l, is often taken to be $\frac{1}{2}$ as determined by (Mualem, 1976). Small changes in pressure head can change the hydraulic conductivity several orders of magnitude; as such, $k(\psi)$ is a highly nonlinear function and this makes Richards equation a highly nonlinear parabolic partial differential equation.

3.3 Implicit sensitivity through matrix-vector products

An efficient solution of the optimization problem requires the derivative of the pressure head at all times, $\boldsymbol{\Psi}$, with respect to the model parameters $\mathbf{m}$, the so called, sensitivity matrix, $\mathbf{J}$. An approximation of the sensitivity matrix can be obtained by using a finite difference method on the forward problem (Šimnek and van Genuchten, 1996; Finsterle and Kowalsky, 2011; Finsterle and Zhang, 2011). One forward problem, or two when using central differences, must be completed for each column in the Jacobian at every iteration of the optimization algorithm. This differentiation has the advantage that it can be applied to any forward problem, however, it is highly inefficient and introduces errors into the inversion that may slow the convergence of the scheme. Automatic differentiation (AD) can also be used (J.~Nocedal and S.~Wright, 1999) however, AD yields a

dense Jacobian and does not take the structure of the problem into consideration. Bitterlich and Knabner (2002) presents three algorithms (finite difference, adjoint, and direct) to directly compute the elements of the dense sensitivity matrix for Richards equation. As we show next, it is possible to explicitly write the derivatives of the Jacobian and evaluate their products with vectors. This is a much more efficient algorithm compared with finite differencing or AD, especially for large scale simulations, since it does not require storing a large and dense matrix and can be used in order to efficiently compute matrix-vector and adjoint matrix vector products. These products can be used for the solution of the Gauss-Newton system when using the conjugate gradient method. The idea has been used extensively for other inverse problems (Haber et al., 2000). The challenge in implementing it for Richards equation lies in differentiating the nonlinear forward problem with respect to the hydraulic parameters. However, as we show next, this can be completed even for highly nonlinear parameterizations, by using implicit differentiation and the chain rule.

The derivative of Richards equation in time involves writing the whole time-stepping process as a block-matrix and taking the derivative. The discrete Richards equation can be written as a function of the model, and for one time-step is written:

$$F(\boldsymbol{\psi}^{n+1}, \boldsymbol{\psi}^n, \mathbf{m}) = \frac{\boldsymbol{\theta}^{n+1}(\boldsymbol{\psi}^{n+1}) - \boldsymbol{\theta}^n(\boldsymbol{\psi}^n)}{\Delta t} + \mathbf{D} \operatorname{diag}\left(\mathbf{k}_{Av}(\boldsymbol{\psi}^{n+1}, \mathbf{m})\right) \mathbf{D}^\top \boldsymbol{\psi}^{n+1} + \mathbf{D}_z \mathbf{k}_{Av}(\boldsymbol{\psi}^{n+1}, \mathbf{m}) = 0$$

(4)

In this case $\mathbf{m}$ contains all the parameters of interest and is a vector. Here we note that $\boldsymbol{\psi}^{n+1}$ and $\boldsymbol{\psi}^n$ are also functions of $\mathbf{m}$. The derivatives of F to the change in the parameters $\mathbf{m}$ can be written in general as

$$\nabla_{\mathbf{m}} F(\boldsymbol{\psi}^n, \boldsymbol{\psi}^{n+1}, \mathbf{m}) = \frac{\partial F}{\partial \mathbf{k}_{Av}} \frac{\partial \mathbf{k}_{Av}}{\partial \mathbf{m}} + \frac{\partial F}{\partial \boldsymbol{\psi}^n} \frac{\partial \boldsymbol{\psi}^n}{\partial \mathbf{m}} + \frac{\partial F}{\partial \boldsymbol{\psi}^{n+1}} \frac{\partial \boldsymbol{\psi}^{n+1}}{\partial \mathbf{m}} = 0 \tag{5}$$

or in more detail

$$\frac{1}{\Delta t} \left(\frac{\partial \boldsymbol{\theta}^{n+1}}{\partial \boldsymbol{\psi}^{n+1}} \frac{\partial \boldsymbol{\psi}^{n+1}}{\partial \mathbf{m}} - \frac{\partial \boldsymbol{\theta}^n}{\partial \boldsymbol{\psi}^n} \frac{\partial \boldsymbol{\psi}^n}{\partial \mathbf{m}} \right) + \mathbf{D} \operatorname{diag}\left(\mathbf{D}^\top \boldsymbol{\psi}^{n+1}\right) \left(\frac{\partial \mathbf{k}_{Av}}{\partial \mathbf{m}} + \frac{\partial \mathbf{k}_{Av}}{\partial \boldsymbol{\psi}^{n+1}} \frac{\partial \boldsymbol{\psi}^{n+1}}{\partial \mathbf{m}} \right)$$

$$+ \mathbf{D} \operatorname{diag}\left(\mathbf{k}_{Av}(\boldsymbol{\psi}^{n+1})\right) \mathbf{D}^\top \frac{\partial \boldsymbol{\psi}^{n+1}}{\partial \mathbf{m}} + \mathbf{D}_z^\top \left(\frac{\partial \mathbf{k}_{Av}}{\partial \mathbf{m}} + \frac{\partial \mathbf{k}_{Av}}{\partial \boldsymbol{\psi}^{n+1}} \frac{\partial \boldsymbol{\psi}^{n+1}}{\partial \mathbf{m}} \right) = 0 \tag{6}$$

The sensitivity matrix is the derivatives of $\boldsymbol{\Psi}$ to $\mathbf{m}$ and the above equation can be viewed as a system of linear equations for its estimation. To solve for $\frac{\partial \boldsymbol{\Psi}}{\partial \mathbf{m}}$ we rearrange the above equation into a block-matrix.

$$\overbrace{\left[-\frac{1}{\Delta t} \frac{\partial \boldsymbol{\theta}^n}{\partial \boldsymbol{\psi}^n} \right]}^{\mathbf{A}_{-1}(\boldsymbol{\psi}^n)} \frac{\partial \boldsymbol{\psi}^n}{\partial \mathbf{m}} + \overbrace{\left[\frac{1}{\Delta t} \frac{\partial \boldsymbol{\theta}^{n+1}}{\partial \boldsymbol{\psi}^{n+1}} + \mathbf{D} \operatorname{diag}\left(\mathbf{D}^\top \boldsymbol{\psi}^{n+1}\right) \frac{\partial \mathbf{k}_{Av}}{\partial \boldsymbol{\psi}^{n+1}} + \mathbf{D} \operatorname{diag}\left(\mathbf{k}_{Av}(\boldsymbol{\psi}^{n+1}, \mathbf{m})\right) \mathbf{D}^\top + \mathbf{D}_z \frac{\partial \mathbf{k}_{Av}}{\partial \boldsymbol{\psi}^{n+1}} \right]}^{\mathbf{A}_0(\boldsymbol{\psi}^{n+1})} \frac{\partial \boldsymbol{\psi}^{n+1}}{\partial \mathbf{m}}$$

$$= \underbrace{\mathbf{D} \operatorname{diag}\left(\mathbf{D}^\top \boldsymbol{\psi}^{n+1}\right) \frac{\partial \mathbf{k}_{Av}}{\partial \mathbf{m}} + \mathbf{D}_z^\top \frac{\partial \mathbf{k}_{Av}}{\partial \mathbf{m}}}_{\mathbf{B}(\psi^{n+1})} \tag{7}$$

Here, we use the subscript notation of $\mathbf{A}_0(\boldsymbol{\psi}^{n+1})$ and $\mathbf{A}_{-1}(\boldsymbol{\psi}^n)$ to represent two diagonals of the large sparse block-matrix $\mathbf{A}(\boldsymbol{\Psi}, \mathbf{m})$. If $\boldsymbol{\psi}_0$ does not depend on the model, meaning the initial

9

conditions are independent, then the block system can be formulated as:

$$
\overbrace{\begin{bmatrix}
\mathbf{A}_0(\psi_1) & & & & \\
\mathbf{A}_{-1}(\psi_1) & \mathbf{A}_0(\psi_2) & & & \\
 & \mathbf{A}_{-1}(\psi_2) & \mathbf{A}_0(\psi_3) & & \\
 & & \ddots & \ddots & \\
 & & & \mathbf{A}_{-1}(\psi_{n_t-1}) & \mathbf{A}_0(\psi_{n_t})
\end{bmatrix}}^{\mathbf{A}(\Psi,\mathbf{m})}
\overbrace{\begin{bmatrix}
\frac{\partial \psi_1}{\partial \mathbf{m}} \\
\frac{\partial \psi_2}{\partial \mathbf{m}} \\
\vdots \\
\frac{\partial \psi_{n_t-1}}{\partial \mathbf{m}} \\
\frac{\partial \psi_{n_t}}{\partial \mathbf{m}}
\end{bmatrix}}^{\frac{\partial \Psi}{\partial \mathbf{m}}}
=
\overbrace{\begin{bmatrix}
\mathbf{B_1}(\psi_1) \\
\mathbf{B_2}(\psi_2) \\
\vdots \\
\mathbf{B_{n-1}}(\psi_{n_t-1}) \\
\mathbf{B_n}(\psi_{n_t})
\end{bmatrix}}^{\mathbf{B}(\Psi,\mathbf{m})}
$$

$$(8)$$

This is a block matrix equation, and both the storage and solving will be expensive if it is explicitly computed. Indeed, its direct computation is equivalent to the adjoint method previously discussed.

However, since only matrix vector products are needed for the inexact Gauss-Newton optimization method the matrix $\mathbf{J}$ is never needed explicitly and only the products of the form $\mathbf{Jv}$ and $\mathbf{J}^\top \mathbf{z}$ are needed, for arbitrary vectors $\mathbf{v}$ and $\mathbf{z}$. Projecting the full sensitivity matrix onto the data-space using $\mathbf{Q}$, results in the following equations for the Jacobian:

$$\mathbf{J} = \mathbf{Q}\mathbf{A}(\Psi,\mathbf{m})^{-1}\mathbf{B}(\Psi,\mathbf{m}) \tag{9a}$$

$$\mathbf{J}^\top = \mathbf{B}(\Psi,\mathbf{m})^\top \mathbf{A}(\Psi,\mathbf{m})^{-\top}\mathbf{Q}^\top \tag{9b}$$

In these equations we are careful to not write $\frac{\partial \Psi}{\partial \mathbf{m}}$, as it is a large dense matrix which we do not want to explicitly compute or store. Additionally, the matrices $\mathbf{A}(\Psi,\mathbf{m})$ and $\mathbf{B}(\Psi,\mathbf{m})$ do not even need to be explicitly formed because the matrix $\mathbf{A}(\Psi,\mathbf{m})$ is a triangular block-system, and can be solved using forward or backward substitution, with only one block-row being solved at a time. To compute the matrix vector product $\mathbf{Jv}$ we use a simple algorithm:

1. Given the vector $\mathbf{v}$ calculate $\mathbf{y} = \mathbf{Bv}$

2. Solve the linear system $\mathbf{Aw} = \mathbf{y}$ for the vector $\mathbf{w}$

3. Set $\mathbf{Jv} = \mathbf{Qw}$

Here we note that steps (1) and (2) are actually done using a for-loop with only one row being computed and stored at a time. As such, only the full solution, Ψ, is stored and all other block-entries may be computed as needed. Similarly, to compute the adjoint $\mathbf{J}^\top \mathbf{z}$ involves the intermediate solve for $\mathbf{y}$ in $\mathbf{A}^\top \mathbf{y} = \mathbf{Q}^\top \mathbf{z}$ and then computation of $\mathbf{B}^\top \mathbf{y}$. Again, the block-matrix solve is done via backward substitution with all block matrix entries being computed as needed. Note that the backward substitution algorithm can be viewed as time stepping, starting from the final time back to the initial time. This is equivalent to the adjoint method that is discussed in Dean and Chen (2011) and reference within. The main difference between our approach and the classical adjoint based method is that our approach yields the *exact* gradient of the discrete system; no such guaranty is given for adjoint based methods.

3.4 Take home points

- Need computational scalability when moving to large 3D distributed parameter estimation.

- Implicit sensitivities are a part of the solution to the scalability.

10

- Parameterizations of the empirically determined hydraulic conductivity function $k(\psi)$:
    - needs reduction of the parameter space to be estimated, and
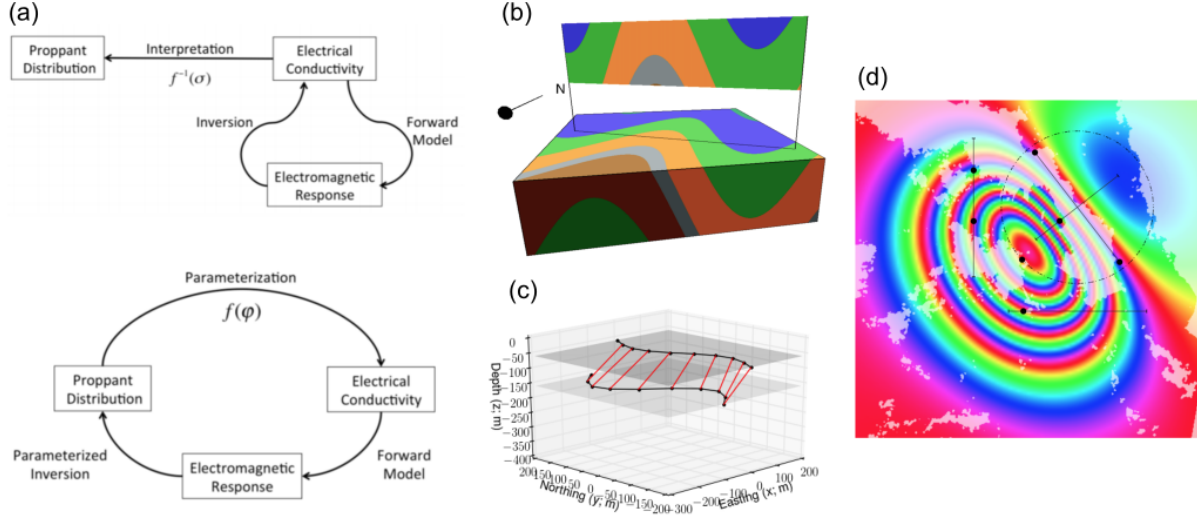   - should be geologically motivated.

Figure 2: Four examples of geologic and a-priori parameterizations: (a) parameterizing a proppant distribution using effective medium theory (Heagy et al., 2014), (b) process based geologic parameterizations (Cockett et al., 2016), (c) interface detection of a seawater intrusion using frequency domain electromagnetics (Kang et al., 2015b), and (d) earthquake parameter estimation using InSAR data and a linear elastic dislocation (Cockett et al., 2014).

4 Interfacing to geologic and a-priori information

Incorporating and quantitatively communicating a-priori and geologic information is a perennial problem in geophysics. There are many existing solutions for inclusion of this information into geophysical inversions, often these methods are completed through parameterizations (e.g. Oldenburg and Li (2005b); Li and Oldenburg (1998, 1996); Pidlisecky et al. (2007, 2011); McMillan et al. (2015)). However, each decision, prediction, and geologic setting is unique: each will often require a new parameterization. I have been involved in a body of work that is researching instances of these parameterizations including: parameterized inversion for proppant distribution in fractures (Heagy et al., 2014), a parameterized model for linear elastic earthquake modelling (Cockett et al., 2014), inverting for seawater intrusion interfaces (Kang et al., 2015a), and a simple geologic parameterization algorithm (Cockett et al., 2016; Cockett, 2013, 2012) (Figure 2). The goal of these works is to understand and abstract these parameterizations into a subset of methods that can generally support the abstract classes of parameterizations. This is an important aspect of an inversion framework: it must have the flexibility to support and interface to arbitrary parameterizations.

A general architecture requires that derivatives be calculated with respect to: (a) multiple physical properties (e.g. electrical conductivity, hydraulic conductivity, magnetic permeability),

(b) the geophysical sources (e.g. waveform, transmitter location), and (c) the receivers (e.g. location, orientation). In order to support the custom parameterizations that are necessary for a unique decision or prediction, this architecture must provide building blocks that are independently extensible. Using examples from multiple fields, including electromagnetics, fluid flow, and parameterized geologic modeling, I plan to investigate the abstractions necessary to simultaneously support a breadth of custom parameterizations.

5   Research questions and thesis outline

Based on the current state of the geoscience inversion community and the observations outlined above, I will arrange my thesis to address three key research topics: (1) developing an extensible framework for geophysical inversions, (2) computationally scalable vadose zone parameter estimation, (3) interfacing to geologic and a-priori information through parametrizations. The overarching goal is to promote integration between the geoscience disciplines; this has guided much of my work. Interdisciplinary integration requires reproducibility, accessibility, and collaboration, as such these are crucial to my work and demonstrated throughout. The research topics that I propose are:

5.1   Simulation and Parameter Estimation Framework

Problem
Geophysical methods for simulation and parameter estimation in geophysics are often treated independently, especially in software, leading to poor interoperability between researchers. This hinders integrations and explorations into combining methods. The terminology and layout of ideas is disorganized internal to and between geophysical and hydrogeologic disciplines, which leads to slow transfer of new ideas. There are no openly available frameworks for geophysical inversions that are independently extensible (i.e. glass-box extensibility).
Question
What is the structure, interconnectivity and architecture of a framework for geophysical inversions, which supports fluid flow, geophysical methods, and geologic integrations?
Approach
Identify and prototype geophysical forward modelling and inverse problems using:

- Finite volume techniques on a variety of meshes:

    - Tensor (1,2,3D); cylindrical; tree (quad/oc); curvilinear

For a variety of test problems including:

- Direct current resistivity

- Time and frequency domain electromagnetics (with Heagy, Kang, Marchant)

- Richards equation

Derive a common framework and ontology from these prototypes. Ensure that the both the ontology and framework are rigorously tested and enforced

- Focus on generalization of the geophysical methodologies

- Identify common pieces and abstract them

- Identify distinct assumptions and separate

Work with collaborators to independently extend (using glass-box extensibility) and refine the framework

- Magnetotellurics (Rosenkjaer)

- Gravity and magnetics (Fournier)

- Induced polarization (Kang)

- Seismic (Smithyman)

An initial version of this framework, called SimPEG (Simulation and Parameter Estimation in Geophysics) has been designed, implemented and published in Computers and Geoscience (Cockett, et al. 2015). There have been over fifteen conference presentations given with varying focuses and applications.

5.2 Implicit calculation of Richards equation sensitivity

Problem
Spatially and temporally extensive data from geophysical imaging methods are increasingly being incorporated with vadose zone fluid simulations to estimate distributions and predict fluid movement. The increase in data constrains the model space such that large parameter estimation problems (large, distributed 2D and 3D estimation problems) can actually be considered. (i.e. a traditional problem would consist of a handful of data points meaning the model space must be further constrained by assumptions, for instance that the earth is 1D, to arrive at a solution). As a result of the increased data, the number of hydraulic parameters to estimate through imaging methods will significantly increase. There are an extensive set of tools for doing forward simulations, however, currently Richards equation inversions are limited by computational scalability. Primarily, this is due to the standard practice of explicit calculation of the sensitivity matrix for the optimization problem.
Question
How do you ensure that the Richards equation inversion is computationally tractable as the number of parameters increases?
Approach

- Leverage the framework development and toolbox created

- Identify the pieces of the sensitivity calculation and tackle the calculation independently and implicitly (rather than explicitly - the current state-of-practice)

- Demonstrate scalability of an implicit sensitivity calculation

- Use a compelling example to communicate these results to the hydrologic audience

There have been a two conference proceedings on earlier versions of this work (Cockett and Haber, 2013a,b). A formalized version of this methodology will be published in June 2016 (targeting Hydrology or Inverse Problems).

5.3 Supporting geologic and a-priori parameterizations

Problem

Geologic and a priori assumptions are often given through parameterizations. There is a body of work on geologic and mathematical parameterizations, however, supporting these in the general case is a difficult proposition. Furthermore, the parameterizations necessary for a specific case study will often need to be unique to answer or predict a specific geoscience hypothesis.

Question

What are the abstractions necessary and the architectural interface to support arbitrary parameterizations in a geophysical inversion?

Approach

- Identify and prototype various types of parameterizations in the different components of the framework, including

    - empirical models (e.g. Van Genuchten)
    - expected distributions of a discretized model (e.g. effective medium theory Heagy et al. (2014))
    - process based geologic models (e.g. Visible Geology Cockett et al. (2016))

- Generalize, classify, and separate the assumptions of the various parameterizations

- Research and comment on the effectiveness of a proposed architectural model for including arbitrary parameterizations

- Demonstrate a geologic, process based parameterization for fluid flow and electromagnetic geophysical methods

The derived architecture to support arbitrary parameterizations and the associated sensitivity calculation, as well as a few key examples of this work, will result in a publication targeting Geophysics or Computers and Geoscience.

5.4 Thesis outline

These research questions will guide the outline of my thesis. Chapter 1 will outline the development of a framework for general geophysical inversions. Chapter 2 will expand on this framework directly for a well-studied problem of electromagnetics and show the implementation and its utility in addressing some unique inversion challenges. Chapter 3 will develop a computationally efficient inversion algorithm for hydraulic parameters in vadose zone flow. Chapter 4 will investigate geologically informed parameterizations and link them to geophysical forward models and inversions.

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., Streicher, O., oct 2013. Astropy: A community python package for astronomy. AAP 558, A33.

Bard, J. B. L., Rhee, S. Y., 2004. Ontologies in biology: design, applications and future challenges. Nature reviews. Genetics 5 (3), 213–222.

Bitterlich, S., Knabner, P., 2002. An efficient method for solving an inverse problem for the Richards equation. Journal of Computational and Applied Mathematics 147, 153–173.

Brooks, R. H., Corey, A. T., 1964. Hydraulic properties of porous media. Hydrology Papers, Colorado State Univ., Fort Collins. 3.

Celia, M. A., Bouloutas, E. T., Zarba, R. L., 1990. A general mass-conservative numerical solution for the unsaturated flow equation. Water Resources Research 26 (7), 1483–1496.
URL http://www.agu.org/pubs/crossref/1990/WR026i007p01483.shtml

Cockett, A. R., Pidlisecky, A., mar 2014. Simulated electrical conductivity response of clogging mechanisms for managed aquifer recharge. GEOPHYSICS 79 (2), D81–D89.
URL http://library.seg.org/doi/abs/10.1190/geo2013-0054.1

Cockett, R., 2012. Visible Geology - Interactive online geologic block modelling. In: AGU Fall Meeting.

Cockett, R., 2013. A Interactive Online Geologic Block Modeling. In: AAPG Hedberg Research Conference - 3D Structural Geologic Interpretation: Earth, Mind, and Machine. AAPG, Reno.

Cockett, R., Funning, G. J., Pratt-Sitaula, B., 2014. Visible Earthquakes : An open visualization and modeling platform for InSAR earthquake data. In: Southern California Earthquake Center - Annual Meeting.

Cockett, R., Haber, E., 2013a. A Numerical Method for Large Scale Estimation of Distributed Hydraulic Conductivity from Richards Equation. In: AGU Fall Meeting.

Cockett, R., Haber, E., 2013b. Inversion for hydraulic conductivity using the unsaturated flow equations. In: SIAM Conference on Mathematical & Computational Issues in the Geosciences.

Cockett, R., Heagy, L. J., Kang, S., Oldenburg, D. W., 2015a. Coupling geophysical terminology and package development. In: SciPy.

Cockett, R., Heagy, L. J., Rosenkjaer, G., Kang, S., 2015b. Development practices and lessons learned in developing SimPEG. In: AGU Fall Meeting.

Cockett, R., Kang, S., Heagy, L. J., Pidlisecky, A., Oldenburg, D. W., 2015c. SimPEG: An open source framework for simulation and gradient based parameter estimation in geophysical applications. Computers & Geosciences 85, 142–154.
URL http://www.sciencedirect.com/science/article/pii/S009830041530056X

Cockett, R., Moran, T., Pidlisecky, A., 2016. Visible Geology: Creative online tools for teaching, learning, and communicating geologic concepts. In: Krantz, R., Ormand, C. J., Freeman, B. (Eds.), Earth, Mind, and Machine: 3D Structural Interpretation, American Association of Petroleum Geologists Hedberg Series. No. 6.

Constable, S. C., Parker, R. L., Constable, C. G., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. Geophysics 52 (3), 289–300.
URL http://link.aip.org/link/?GPY/52/289/1

Daily, W., Ramirez, A., LaBrecque, D., Nitao, J., 1992. Electrical resistivity tomography of vadose water movement. Water Resources Research 28 (5), 1429–1442.
URL http://www.agu.org/pubs/crossref/1992/91WR03087.shtml

Dean, O. S., Chen, Y., 2011. Recent progress on reservoir history matching: a review. Comput Geoscience 15, 185–221.

Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S. A., Green, A. G., nov 2010. Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. GEOPHYSICS 75 (6), G53–G64.
URL http://library.seg.org/doi/abs/10.1190/1.3496476

Finsterle, S., Kowalsky, M. B., jun 2011. A truncated LevenbergMarquardt algorithm for the calibration of highly parameterized nonlinear models. Computers & Geosciences 37 (6), 731–738.
URL http://linkinghub.elsevier.com/retrieve/pii/S0098300410003729

Finsterle, S., Zhang, Y., jul 2011. Solving iTOUGH2 simulation and optimization problems using the PEST protocol. Environmental Modelling & Software 26 (7), 959–968.
URL http://linkinghub.elsevier.com/retrieve/pii/S1364815211000521

Fohring, J., Cockett, R., Haber, E., 2012. A multiphysics approach to electromagnetic inversion with application to enhanced oil recovery. In: AGU Fall Meeting.

Gao, G., Abubakar, A., Habashy, T. M., may 2012. Joint petrophysical inversion of electromagnetic and full-waveform seismic data. GEOPHYSICS 77 (3), WA3–WA18.
URL http://library.seg.org/doi/abs/10.1190/geo2011-0157.1

Haber, E., Ascher, U., Oldenburg, D., 2000. On Optimization Techniques for Solving Nonlinear Inverse Problems. Inverse problems 16, 1263–1280.

Haber, E., Oldenburg, D., 1997. Joint Inversion A Structural Approach. Inverse Problems 13, 63–67.

Hansen, T. M., Cordua, K. S., Looms, M. C., Mosegaard, K., mar 2013. SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information. Computers & Geosciences 52, 481–492.
URL http://linkinghub.elsevier.com/retrieve/pii/S0098300412003408

Harbaugh, A., 2005. The u.s. geological survey modular ground-water model the ground-water flow
    process. U.S. Geological Survey Techniques and Methods (6), A16.

Heagy, L. J., Cockett, A. R., Oldenburg, D. W., aug 2014. Parametrized inversion framework for
    proppant volume in a hydraulically fractured reservoir. In: SEG Technical Program Expanded
    Abstracts 2014. Society of Exploration Geophysicists, pp. 865–869.
    URL http://library.seg.org/doi/abs/10.1190/segam2014-1639.1

Heagy, L. J., Cockett, R., Kang, S., Oldenburg, D. W., 2015a. Real simulation tools in introductory
    courses: packaging and repurposing our research code. In: AGU Fall Meeting.

Heagy, L. J., Cockett, R., Oldenburg, D. W., 2015b. Using Python to span the gap between
    education, research, and industry applications in geophysics. In: SciPy.

Heagy, L. J., Cockett, R., Oldenburg, D. W., Wilt, M., aug 2015c. Modelling electromagnetic
    problems in the presence of cased wells. In: SEG Technical Program Expanded Abstracts 2015.
    Society of Exploration Geophysicists, pp. 699–703.
    URL http://library.seg.org/doi/10.1190/segam2015-5931035.1

Hewett, R., Demanet, L., the PySIT Team, 2013. PySIT: Python seismic imaging toolbox v0.5.
    Release 0.5.
    URL http://www.pysit.org

Holscher, E., Leifer, C., Grace, B., 2010. Read the docs.
    URL https://docs.readthedocs.org

J.~Nocedal, S.~Wright, 1999. Numerical Optimization. Springer, New York.

Jones, E., Oliphant, T., Peterson, P., Others, 2001. {SciPy}: Open source scientific tools for
    {Python}.
    URL http://www.scipy.org/

Kalderimis, J., Meyer, M., 2011. Travis ci gmbh.
    URL https://travis-ci.org

Kang, S., Cockett, R., Heagy, L. J., 2014. Moving between dimensions in electromagnetic inversions
    with a consistent framework. In: AGU Fall Meeting.

Kang, S., Cockett, R., Heagy, L. J., Oldenburg, D. W., aug 2015a. Moving between dimensions in
    electromagnetic inversions. In: SEG Technical Program Expanded Abstracts 2015. No. 2. Society
    of Exploration Geophysicists, pp. 5000–5004.
    URL http://library.seg.org/doi/10.1190/segam2015-5930379.1

Kang, S., Cockett, R., Heagy, L. J., Oldenburg, D. W., 2015b. Moving between dimensions in
    electromagnetic inversions (2).

Kang, S., Oldenburg, D. W., 2016. On recovering distributed IP information from inductive source
    time domain electromagnetic data (In Review). Geophysical Journal International.

Kelbert, A., Meqbel, N., Egbert, G. D., Tandon, K., May 2014. ModEM: A modular system for inversion of electromagnetic geophysical data. Computers & Geosciences 66, 40–53.
URL http://linkinghub.elsevier.com/retrieve/pii/S0098300414000211

Lelièvre, P. G., Oldenburg, D. W., Williams, N. C., 2009. Integrating geological and geophysical data through advanced constrained inversions. Exploration Geophysics 40 (4), 334–341.

Li, Y., Key, K., 2007. 2d marine controlled-source electromagnetic modeling: Part 1 an adaptive finite-element algorithm. GEOPHYSICS 72 (2), WA51–WA62.
URL http://dx.doi.org/10.1190/1.2432262

Li, Y., Oldenburg, D. W., 1996. 3D Inversion of magnetic data. Geophysics 61, 394–408.

Li, Y., Oldenburg, D. W., jan 1998. 3-D inversion of gravity data. Geophysics 63 (1), 109–119.
URL http://library.seg.org/doi/abs/10.1190/1.1444302

Li, Y., Oldenburg, D. W., jan 2000. Incorporating geological dip information into geophysical inversions. Geophysics 65 (1), 148–157.
URL http://library.seg.org/doi/abs/10.1190/1.1444705

Liang, W.-L., Uchida, T., mar 2014. Effects of topography and soil depth on saturated-zone dynamics in steep hillslopes explored using the three-dimensional Richards equation. Journal of Hydrology 510, 124–136.
URL http://linkinghub.elsevier.com/retrieve/pii/S0022169413009268

McMillan, M. S., Schwarzbach, C., Haber, E., Oldenburg, D. W., 2015. 3D parametric hybrid inversion of time-domain airborne electromagnetic data. Geophysics 80 (6), K25–K36.
URL http://library.seg.org/doi/10.1190/geo2015-0141.1

Merwin, N., Donahoe, L., Mcangus, E., 2015. Coveralls.
URL https://coveralls.io

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research 12 (3), 513.
URL http://www.agu.org/pubs/crossref/1976/WR012i003p00513.shtml

Oldenburg, D. W., Li, Y., 2005a. 5. Inversion for Applied Geophysics: A Tutorial. Ch. 5, pp. 89–150.
URL http://library.seg.org/doi/abs/10.1190/1.9781560801719.ch5

Oldenburg, D. W., Li, Y., 2005b. Inversion for applied geophysics: A Tutorial. Investigations in geophysics (13), 89–150.

Orgogozo, L., Renon, N., Soulaine, C., Hénon, F., Tomer, S., Labat, D., Pokrovsky, O., Sekhar, M., Ababou, R., Quintard, M., dec 2014. An open source massively parallel solver for Richards equation: Mechanistic modelling of water fluxes at the watershed scale. Computer Physics Communications 185 (12), 3358–3371.
URL http://linkinghub.elsevier.com/retrieve/pii/S0010465514002719

Park, S., 1998. Fluid migration in the vadose zone from 3-D inversion of resistivity monitoring data. Geophysics 63 (1), 41–51.
URL http://link.aip.org/link/GPYSA7/v63/i1/p41/s1{&}Agg=doi

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in {P}ython. Journal of Machine Learning Research 12, 2825–2830.

Pidlisecky, A., Cockett, a. R., Knight, R., 2013. Electrical Conductivity Probes for Studying Vadose Zone Processes: Advances in Data Acquisition and Analysis. Vadose Zone Journal 12 (1), 1–12.
URL https://www.soils.org/publications/vzj/abstracts/12/1/vzj2012.0073

Pidlisecky, A., Haber, E., Knight, R., 2007. RESINVM3D : A 3D resistivity inversion package. Geophysics 72 (2), H1–H10.

Pidlisecky, A., Singha, K., Day-Lewis, F. D., 2011. A distribution-based parametrization for improved tomographic imaging of solute plumes. Geophysical Journal International 187, 214–224.

Racz, A. J., Fisher, A. T., Schmidt, C. M., Lockwood, B. S., Huertos, M. L., nov 2011. Spatial and Temporal Infiltration Dynamics During Managed Aquifer Recharge. Ground water, 1–9.
URL http://www.ncbi.nlm.nih.gov/pubmed/22050208

Rosenkjaer, G., Heagy, L. J., Cockett, R., 2015a. Practical integration of processing, inversion and visualization of magnetotelluric geophysical data. In: SciPy.

Rosenkjaer, G., Heagy, L. J., Cockett, R., Kang, S., Oldenburg, D., 2015b. Using simpegMT to demonstrate a modular modelling and inversion framework applied in a workflow for a hydrothermal system. In: AGU Fall Meeting.

Šimnek, J., van Genuchten, M. T., 1996. Estimating Unsaturated Soil Hydraulic Properties from Tension Disc Infiltrometer Data by Numerical Inversion. Water Resources Research 32 (9), 2683.
URL http://www.agu.org/pubs/crossref/1996/96WR01525.shtml

Steklova, K., Haber, E., Cockett, R., 2012. Seawater intrusions: Coupling groundwater model and geophysical data. In: AGU Fall Meeting.

Uieda, L., Jr, V. C. O., Ferreira, A., dos Santos, H. B., Jr., J. F. C., 2014. Fatiando a Terra: a Python package for modeling and inversion in geophysics.
URL http://dx.doi.org/10.6084/m9.figshare.1115194

van Genuchten, M. T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Science Society of America Journal 44 (5), 892.
URL https://www.soils.org/publications/sssaj/abstracts/44/5/SS0440050892

Wilson, G., Aruliah, D. a., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., Wilson, P., jan 2014. Best practices for scientific computing. PLoS biology 12 (1), e1001745.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3886731{&}tool=pmcentrez{&}rendertype=abstract