A COMPARISON OF SEQUENTIAL AND INTEGRATED DATA FUSION FOR ESTIMATING HYDROLOGIC PROPERTIES DURING A SYNTHETIC GPR MONITORED INFILTRATION EVENT

Guy Sicilia jr
Clemson University, tom@coffeepot.org

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Part of the Hydrology Commons

Recommended Citation
https://tigerprints.clemson.edu/all_theses/463

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
A COMPARISON OF SEQUENTIAL AND INTEGRATED DATA FUSION FOR
ESTIMATING HYDROLOGIC PROPERTIES DURING A
SYNTHETIC GPR MONITORED
INFILTRATION EVENT

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Hydrogeology

by
Guy Thomas Sicilia Jr
August 2008

Accepted by:
Stephen Moysey, Committee Chair
Ron Falta
Larry Murdoch
ABSTRACT

Constraining parameters that govern variably saturated flow is important for applications ranging from quantifying water availability for ecosystems to constraining recharge rates and contaminant fluxes to groundwater. In this study I explore the effectiveness of sequential versus integrated data fusion for estimating unsaturated flow parameters using ground penetrating radar (GPR) data. In Sequential Data Fusion (SDF), geophysical imaging is used to create a map of the geophysical properties of the subsurface. These properties are then transformed to hydrologic properties that can be used to constrain an independent hydrologic inverse problem. In contrast, Integrated Data Fusion (IDF) uses the geophysical data to directly constrain hydrologic properties of interest without performing the intermediate geophysical imaging step. The comparison of SDF and IDF is performed for a synthetic study of 2D infiltration into a homogeneous soil from a constant flux point source located at the ground surface. The focus is on results for the estimation of intrinsic permeability (k) from cross-borehole GPR traveltimes collected throughout the duration of the infiltration event. The target permeability (k=7.4x10^{-12}m^2) is uniform over the 20 meter by 20 meter area modeled in this study; though the soil is homogeneous, water content is both spatially variable and transient. I use TOUGH2 to simulate infiltration, MATLAB to simulate GPR traveltimes, and PEST to perform the parameter estimation. To quantitatively compare SDF and IDF, I calculate the normalized error in estimated permeability for each method. In my study, I investigated the performance of the data fusion methods under varying survey geometries
by changing the antenna spacing. In all cases I have found that IDF significantly outperforms SDF. For large antenna separations (1.7-6.7m) SDF produces an average error in estimated permeability of 73% while IDF errors are only 6%. As ray density is increased for antenna separations of 1.0-1.5m, average estimation error for SDF drops to 72%, but is reduced to only 3% for IDF. Also, SDF estimates are consistently biased lower than the target value, while IDF results are unbiased. My results suggest the IDF is a powerful new approach for hydrologic characterization of the subsurface using geophysical measurements.
DEDICATION

For Cynthia and Julia, for all their patience.
ACKNOWLEDGMENTS

I would like to acknowledge the Moysey Research Center members for all of their support, Stephen Moysey, for his inspiration, my committee members for their guidance, and Robert Mitchell for setting me on the path.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>HYDROGEOLOGIC BACKGROUND</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Overview</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Unsaturated flow</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Theory of flow in unsaturated materials</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Estimation of parameters for use in unsaturated flow models</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Geophysical monitoring of unsaturated flow</td>
<td>13</td>
</tr>
<tr>
<td>III</td>
<td>GROUND PENETRATING RADAR</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Fundamentals of GPR</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>GPR Forward models</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Geophysical inversion</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Borehole GPR Tomography example</td>
<td>24</td>
</tr>
<tr>
<td>IV</td>
<td>HYDROLOGIC MODEL CALIBRATION USING GEOPHYSICAL DATA</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Approaches to using geophysical measurements in hydrologic</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inverse problems</td>
<td>31</td>
</tr>
</tbody>
</table>
### Table of Contents (Continued)

#### V. SYNTHETIC STUDIES OF BGPR-MONITORED INFILTRATION ......36
- Setup of reference model used in synthetic studies ...............................36
- Estimation of permeability using sequential data fusion .......................39
- SDF results.............................................................................................41
- Moment analysis ....................................................................................46
- Estimation of permeability using integrated data fusion .......................47
- IDF results..............................................................................................48
- Comparison of SDF and IDF .................................................................50
- Estimation of Permeability with unknown porosity ..............................53
- Impact of noise on permeability estimation...........................................56

#### VI. CONCLUSIONS..........................................................................................59
- Recommendations for future work ........................................................60

**APPENDICES ...............................................................................................................61**
- A: List of recorded time steps.................................................................62
- B: Digital Appendix..................................................................................63

**REFERENCES ..............................................................................................................64**
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Soil parameter values used in simulations</td>
<td>37</td>
</tr>
<tr>
<td>5.2</td>
<td>SDF Estimates of intrinsic permeability (m²) with varying ray density</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>and using inversion method based on least squares</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>SDF Estimates of intrinsic permeability (m²) with varying ray density</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>and using inversion which fit conceptual model</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>IDF Estimates of intrinsic permeability (m²) with varying ray density</td>
<td>49</td>
</tr>
<tr>
<td>5.5</td>
<td>Results of adding Gaussian error to traveltimes</td>
<td>57</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>15</td>
</tr>
<tr>
<td>3.1</td>
<td>17</td>
</tr>
<tr>
<td>3.2</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>24</td>
</tr>
<tr>
<td>3.4</td>
<td>25</td>
</tr>
<tr>
<td>3.5</td>
<td>27</td>
</tr>
<tr>
<td>3.6</td>
<td>28</td>
</tr>
<tr>
<td>3.7</td>
<td>29</td>
</tr>
<tr>
<td>4.1</td>
<td>31</td>
</tr>
<tr>
<td>4.2</td>
<td>33</td>
</tr>
<tr>
<td>4.3</td>
<td>34</td>
</tr>
<tr>
<td>4.4</td>
<td>34</td>
</tr>
</tbody>
</table>

2.1 Comparison of van Genuchten and Brooks-Corey desaturation curves for a fine sand \( m=0.871, \alpha=0.202, \Psi^d=41\text{cm}, \lambda=3.7 \) (Charbeneau, 2000)

2.2 Conceptual model of three subsurface systems being perturbed by adding a water slug.

2.3 Example of one way geophysical methods can be used to calibrate hydrogeologic models.

3.1 Conceptual model of reflected, refracted, and transmitted waves (wave path represented by a ray)

3.2 Schematic of typical GPR borehole arrays (Huismann et al., 2003)

3.3 Box with slowness 20 in the center of a matrix with slowness of 1

3.4 400 GPR rays in a MOG array

3.5 Inverted GPR traveltimes with slowness weighting

3.6 Inverted GPR data with continuity weighting

3.7 Inverted GPR data with continuity and slowness weighting

4.1 Flow chart describing sequential data fusion (Moysey et. Al. 2006)

4.2 Flow chart describing integrated data fusion (Moysey et. Al. 2006)

4.3 Schematic representations of wetting front locations relative to zero-offset profiling (ZOP) borehole ground penetrating radar (BGPR) antennae and associated BGPR first arriving travel paths.

4.4 Error surface of van Genuchten parameter (Rucker and Ferre, 2004)
List of Figures (Continued)

4.5 the left plots are traveltime datasets from a ZOP survey fit to a one layer infiltration model, while the right two plots are from a five layer model (Looms et al., 2008a) ................................................................. 35

5.1 Infiltration plume image every third recorded time First row left to right: 3100 seconds, 133 days, 417 days; second row left to right: 4.78 years, 6.34 years, 7.89 years; third row left to right: 9.44 years, 10.49 years, 13.4 years. Magnitudes are water content values. .......... 37

5.2 Flow chart describing sequential data fusion as used in this study........... 40

5.3 Inversion of the synthetic field data using the MSE method of selecting filter weights. Magnitude is water content........................................ 42

5.4 Scaled objective functions from SDF estimations using varying ray density and MSE weighting (for each case the sse is normalized by the maximum SSE). ................................................................. 43

5.5 Inverted Infiltration plume image every third recorded time. Magnitudes are water content values. First row left to right: 3100 seconds, 133 days, 417 days; second row left to right: 4.78 years, 6.34 years, 7.89 years; third row left to right: 9.44 years, 10.49 years, 13.4 years........ 45

5.6 Results of spatial moment analysis.......................................................... 47

5.7 Flow chart describing Integrated Data Fusion as used in this study........... 48

5.8 Scaled objective functions from IDF estimations using varying ray density................................................................................................. 50

5.9 Scaled objective functions from IDF estimations using varying ray density................................................................................................. 51

5.10 Log scale scaled objective functions for SDF and IDF 400 ray estimates .. 52
List of Figures (Continued)

5.11 Variation in SDF and IDF intrinsic permeability estimates caused by varying GPR ray density ................................................................. 53

5.12 3D error surface comparing RMSE for porosity and intrinsic permeability combinations from the tough2 output .............................. 54

5.13 3D error surface comparing RMSE for porosity and intrinsic permeability combinations using the SDF method ............................. 55

5.14 3D error surface comparing RMSE for porosity and intrinsic permeability combinations using the SDF method ............................. 56

5.15 Permeability estimates with increasing Gaussian noise added to GPR Data .................................................................................................................. 57

5.16 Error surfaces of RMSE water contents created using the SDF method and adding increasing signal noise ........................................ 58

5.17 Error surfaces of RMSE traveltimes created using the IDF method And increasing signal noise ................................................................. 58
CHAPTER ONE
INTRODUCTION

The ability to estimate soil properties is critical for modeling infiltration, contaminant transport, and aquifer recharge (USNCRM, 2001). However, conventional tests for the characterization of unsaturated flow parameters are both invasive and local, making them inappropriate for understanding field-scale flow and transport processes in heterogeneous media. In contrast, geophysical methods provide the possibility to non-invasively study the vadose zone over large spatial extents.

This paper focuses on improved ways to quantify unsaturated flow parameters of soils using geophysical methods, specifically Ground Penetrating Radar (GPR). GPR has been used to map spatial and temporal variations of water content over volumes ranging from tens of centimeters to meters (Davis and Annan, 1989; Huismann et al., 2003). Conceptually, these indirect measurements of water content could be used to improve the calibration of unsaturated flow models, but operationally they typically fall short. That is, GPR-based water content estimates are frequently qualitative rather than quantitative due to the limitations of geophysical imaging and inversion techniques (Moysey and Knight, 2004; Moysey et al., 2005). As a result, it is questionable whether GPR used in this sequential method can provide estimates of water content that are sufficiently accurate to act as effective constraints for hydrologic parameter estimation problems.

A promising research thread is investigating whether geophysical measurements can be directly linked to hydrologic parameters. The hypothesis of this research is that using coupled geophysical and hydrologic forward models can overcome the resolution
and non-uniqueness limitations of geophysical inversion that have historically hindered the use of GPR (Kowalsky et al., 2005; Lambot et al., 2004; Moysey et al., 2006; Rucker and Ferre, 2004; Vereecken et al., 2007). For example, varying the parameters of an unsaturated flow model affects the predicted distribution of water content in the subsurface, thereby leading to changes in the dielectric constant distribution. Because the dielectric constant of the subsurface controls the velocity at which an electromagnetic (EM) wave can travel through it, the initial perturbation in hydrologic parameters ultimately leads to a change in the predicted EM traveltimes between two boreholes observed with GPR. Sensitivity of GPR traveltimes to the hydrologic parameters controlling flow implies that hydrologic characterization is possible without the need to first estimate water contents from the GPR traveltime data, thereby avoiding the geophysical inversion step and its associated problems. To date, however, no study has compared the limitations or benefits of sequential versus coupled approaches to data integration for GPR.

The underlying objective of this thesis is to investigate the merits of sequential inversion versus coupled methods for estimating intrinsic permeability with cross-borehole GPR traveltime tomography. Intrinsic permeability is the target parameter in this study because a field scale value is not easily measurable, it is an important property in unsaturated flow, and it can be estimated in unsaturated conditions. I hypothesize that relative to coupled methods, sequential methods will provide poor predictions of the intrinsic permeability values because the geophysical inversion of GPR traveltimes can introduce a lower bias to water content estimates. I also hypothesize that sequential
methods will be more sensitive to the number of GPR measurements because as data becomes more scarce, more weight is put on criteria that are added to the inversion of the traveltime data. Coupled models will still be sensitive to data error, but because fewer model parameters need to be estimated, the sensitivity will not be amplified.

To accomplish the comparison of inversion methods, I simulated an infiltration event monitored by a GPR survey. I then estimate intrinsic permeability using both data integration methods, with varying data density and survey designs. Chapter two will give background on unsaturated flow, the parameters that govern it, and the methods that have been used on the lab and field scales to assign values to these parameters. Chapter three is an overview of geophysical methods, specifically GPR and its utility for use in hydrologic problems. Chapter four is a discussion of model calibration using inverse methods. Chapter five contains the details of the models used to accomplish the goal of this research, and the presentation of results. Chapter six presents the conclusions and discussion of the results.
2.1 Overview

The purpose of this research is to compare two methods of estimating intrinsic permeability using geophysical methods. This chapter provides background on the processes, measurements, and estimation techniques involved in characterizing unsaturated flows using both hydrologic and geophysical approaches. First, the parameters that control unsaturated flow and how these parameters have been estimated at lab and field scales in traditional hydrologic settings will be discussed. I will then expand on geophysical methods for estimating water content, which requires discussion of the inversion techniques used in imaging problems. Finally, I will discuss how geophysical methods have been used as constraints in hydrologic estimation problems.

2.2 Unsaturated flow

Infiltration of liquids into the subsurface has wide ranging implications, from determining aquifer recharge to influencing contaminant transport and plume migration rates in the vadose zone. This section will first introduce the parameters of import, and then examine the common laboratory and field methods traditionally used to determine values for these parameters. A discussion of geophysical methods as applied to hydrologic problems and calibration of models complete this section and chapter.
2.3 Theory of flow in unsaturated materials

Downward flow in the vadose zone is controlled by soil unsaturated hydraulic conductivity $K(\psi)$ and pressure head as described by Richards equation (Richards, 1931).

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}[K(\psi)(\frac{\partial \psi}{\partial z} + 1)]$$

(1)
given $z$ is the elevation above a vertical datum, $\theta$ is the water content, $t$ is time, and $\psi$ is suction head. Hydraulic conductivity, $(K)$ is defined as:

$$K = \frac{k \rho g}{\mu},$$

(2)

where $k$ is the medium’s intrinsic permeability, $g$ is acceleration due to gravity, $\rho$ is the density of the fluid, and $\mu$ is the viscosity of the fluid.

Hydraulic conductivity is often viewed as the most important parameter for flow and transport problems (Jabro and Evans, 2006; Mohanty et al., 1994). However, as we see in equation 2, hydraulic conductivity includes a more basic property, that is intrinsic permeability. In variably saturated materials, intrinsic permeability, and therefore hydraulic conductivity are dependant on water content (Fetter, 2001). If a second phase (air in an unsaturated case) is added to the system, the permeability becomes a function of the saturation.

$$K = k_{sat}k(\theta)\rho g/\mu$$

(3)

$k_{sat}$ is the saturated intrinsic permeability and $k(\theta)$ ranges from 0 to 1.

This results in a maximum intrinsic permeability value and corresponding hydraulic conductivity value when the soil is saturated. One of the most commonly used models to
describe the dependence of water content on pressure is the Van Genuchten-Mualem model (Van Genuchten, 1980).

\[
S^* = ((1 + \alpha \Psi)^n)^{-m}
\]  

(4)

where \( S^* = \frac{\theta - \theta_{ir}}{\theta_{sat} - \theta_{ir}} \), \( \theta \) is water content, \( \theta_{sat} \) is saturated water content, \( \theta_{ir} \) is the irreducible water content, \( \Psi \) is the capillary pressure head and \( m = 1 - (1/n) \). \( S^* \) is the effective saturation of the material. The variable \( \alpha \) is related to the inverse of the air entry, or bubbling pressure and \( n \) reflects the pore size distribution of the material.

Another commonly used model for the pressure-saturation curve is the Brooks-Corey (Brooks and Corey, 1964) equation:

\[
S^* = \begin{cases} 
1, & \Psi \leq \Psi_b \\
\left(\frac{\Psi}{\Psi_b}\right)\lambda, & \Psi > \Psi_b 
\end{cases}
\]  

(5)

\[
S^* = \frac{\theta - \theta_{ir}}{n - \theta_{ir}} 
\]

where \( \Psi \) is the suction head, \( \Psi_b \) is the air entry or bubbling pressure, \( \theta \) is water content and \( \lambda \) is a fitting parameter.

The van Genuchten model is fully differentiable (Charbeneau, 2000), and closely fits laboratory desaturation experiments (Van Genuchten, 1980). At the air entry pressure of the medium, \( \Psi = \Psi_b \), the derivative of the Brooks-Corey model is discontinuous (Figure 2.1), causing potential instability in numerical models (Charbeneau, 2000). Russo (1988)
showed that the van Genuchten model can more accurately predict infiltration data than those of Brooks-Corey or Gardner when used to estimate hydraulic properties (Russo, 1988).

In Figure 2.1, we see that as water content increases to saturation, suction head drops; causing intrinsic permeability and hydraulic conductivity to approach their maximum magnitudes.

![Comparison of van Genuchten and Brooks-Corey desaturation curves for a fine sand](image)

**Figure 2.1:** Comparison of van Genuchten and Brooks-Corey desaturation curves for a fine sand

A soil property that does not govern unsaturated flow directly, but does affect how much liquid the soil can hold before becoming saturated is porosity. Porosity is defined as volume of voids/ total volume, and typically ranges from 20-58% in soils (Marshall et al., 1996), and is usually measured in the lab gravimetrically (Charbeneau,
2.4 Estimation of parameters for use in unsaturated flow models

Models based on Richard’s Equation are able to represent unsaturated flow, but these models are only as accurate as the soil parameter values that are input into them (Mace et al., 1998). Accurate in situ estimates of intrinsic permeability (k), porosity (n), and model-specific fitting parameters (e.g., α, n, λ) would increase model accuracy. However, field-scale values for k and n are not easily obtained as the subsurface is never homogeneous and laboratory or field methods have limited areas of influence (Rubin and Hubbard, 2005; Sudicky and Huyakorn, 1991; Zheng and Gorelick, 2003).

The next section will discuss commonly used methods of estimating or calculating values for permeability, porosity, and capillary curves. First, laboratory methods are discussed, then methods that are applicable to a field site, and finally geophysical methods are introduced.

Intrinsic permeability

In the lab, intrinsic permeability can be measured using a permeameter. A permeameter works by allowing flow at through a soil-filled column under a known head gradient. By measuring the discharge, Darcy’s Law (Darcy 1856) (6) can be used to determine the hydraulic conductivity of the sample.
\[ Q = -KA \left( \frac{dh}{dx} \right) \]  

(6)

Where \( Q \) is the volumetric discharge, \( A \) is the cross sectional area of the core, \( \frac{dh}{dx} \) is the head gradient, and \( K \) is the hydraulic conductivity.

This method is susceptible to a variety of errors. Cores may have been disturbed during collection or transport, may not have been representative of a heterogeneous field site, or may have a disturbed skin around the outer edges causing a preferential flow path and a higher value of hydraulic conductivity. This skin effect is often mitigated by coating the outer edge of the column in wax (Kool et al., 1985), or using a flexible wall permeameter. This method provides a local measure of hydraulic conductivity, and therefore does not represent field-scale heterogeneous systems with preferential flow paths (i.e. around clay lenses).

**Porosity**

Porosity is measured with a gravimetric method in the lab by submerging a soil sample with a known volume in water until it is saturated, weighing the sample, oven drying the sample, and then reweighing to determine the volume of water lost. This gravimetric method is sensitive to organic content in the sample. As organic materials dry, they often lead to an overprediction of porosity (Van Genuchten et al., 1999). Alternatively, an oven-dried sample may be exposed to a pressurized gas. The amount of gas intruding into the sample is measured, and the volume of pores of a size that relates to
that specific pressure is calculated. The pressure is then changed to a new constant, and the volumes of corresponding pores are recorded. The cumulative volume of these pores is the porosity.

*Water retention measurements*

Saturation curves are created by placing a container full of saturated soil on a permeable ceramic plate and applying a vacuum. As the pressure is increased, the cumulative amount of water that has passed through the ceramic plate is measured. Since the volume of soil is known, the water content can be determined as a function of pressure (Charbeneau, 2000).

These methods are quantifiable and controllable, but laboratory tests by definition involve disturbing the samples (Cazaux and Didier, 2002), so results may not be representative of the true values. As such, properties determined using laboratory methods may not be accurate representations of field-scale soils (Warrick, 1993). To avoid these issues of scale and sample integrity, it is often more desirable to measure or estimate soil properties with field-scale methods.

An inherent problem with field-scale measurements is that much of the volume of interest is inaccessible for detailed characterization (Won, 1990). In addition, behavior of the subsurface is different from the lab due to preferential pathways and heterogeneity. As a result, most measurement techniques in the field involve perturbing a system and measuring the response.
Intrinsic permeability in the vadose zone on the field scale is often estimated using Guelph Permeameters. These devices measure the rate of water flowing into a shallow hole at a constant head. Using this information, it is possible to calculate hydraulic conductivity and soil sorptivity (the ability of a soil to absorb water through capillary forces) for an area on the decimeter scale. Capillary pressure curves can be created using data from a set of tensiometers, which measure a soil suction pressure at varying depths (Stannard, 1986). Guelph Permeameters and tensiometers provide point measurements of permeability and capillary pressures (Stephens and Newman, 1982), but require many time consuming measurements to describe a large study area.

The above methods can provide estimates for soil properties in the area close to the measurement device or well. The perturbation usually involves introducing or removing a fluid, water or air, into the subsurface, so the area of measurement is dependant on the volume introduced, length of time in which the fluid is added or removed, and the rate at which the perturbation propagates through the medium. Heterogeneities may cause values for intrinsic permeability which are not average field scale values, but rather apparent permeabilities.
Figure 2.2 conceptual model of three subsurface systems being perturbed by adding a volume of water.

Figure 2.2 illustrates this concept for a Guelph permeameter test in three soils. The wells are in (from left to right) a homogeneous clay, homogeneous sand, and sand underlain by a clay lens. When the test is initiated, a volume of water enters the subsurface. For the first two homogeneous cases, the measured values will be an adequate field value. However, the third case will result in a value of hydraulic conductivity (and corresponding intrinsic permeability) that is an average value for the system. The dependence of fluid flow on heterogeneity either requires that many localized measurements are performed to characterize a site or assume that the measured value is representative.

However, if the area of interest is large and time is limited, methods exist which can provide information about the subsurface by using indirect methods, such as...
geophysics. In many cases, this allows heterogeneity in the subsurface to be constrained more readily than by hydrologic tests alone.

2.5 Geophysical Monitoring of Unsaturated Flow

Geophysical methods can be employed to provide field-scale measurements of subsurface properties. Geophysical measurements are indirect, meaning the data that is measured is not the parameter of interest, but is rather correlated to that parameter in some way. This allows the water content to be indirectly measured without excessive disturbance of the subsurface.

Many geophysical measurements are sensitive to water content (Huismann et al., 2003), making them good candidates for monitoring infiltration (Binley et al., 2001; Lim et al., 1989) and calibrating hydrologic models (Bowling et al., 2005). For example, time domain reflectometry uses traveltimes of reflected electromagnetic pulses to determine dielectric constants. These dielectric constants are then correlated to an average water content over the length of the device (Topp and Davis, 1985). This method can produce horizontal or vertical measurements over time, but devices are typically only 8 inches to 2 feet in length, so field scale values are not feasible. Neutron probes can provide average values at a localized point, but involve complying with extensive regulations for their use (Fayer, 2005). Electrical resistivity tomography (ERT) can be used to estimate soil types and infer formation thicknesses (Binley and Kemna, 2005), but have poor spatial resolution due to sampling density limitations, and are highly sensitive to clay layers or salinity. Electromagnetic induction uses very low voltage and amperage
currents in a loop which surrounds the study area to induce a magnetic field in water bearing regions and map water bearing aquifers (Willowstick, 2005), but requires that there are no power lines or metal objects in the study area, requires a very large wire loop to be deployed, and has limited resolution in near surface applications. Surface and borehole ground penetrating radar offer an advantage over these methods because GPR has applications from lab to field scale, is versatile since there are numerous survey geometries, and has often been used to measure water content (Huismann et al., 2003). The time required to complete a GPR tomographic survey is generally on the time-scale of hours, making GPR ideal for the dynamic monitoring of infiltration.

Geophysical methods have been used to monitor groundwater leachate concentrations (Tweeton et al., 1991), seasonal soil moisture variability (French and Binley, 2004; French et al., 2002), and as a tool for improved agricultural practices (Michot et al., 2003). This ability to monitor hydrologic processes makes geophysical methods additionally useful for calibrating hydrologic models.

One method of calibrating hydrologic models is using geophysical methods to monitor field scale water content and trying to create a hydrologic model that matches the water content behavior observed geophysically (Deiana et al., 2008). A hydrologic model is created based on the conceptual model at the site and the parameters of the model are varied. When the modeled water contents closely match the geophysical measurements, the model parameter values are potentially viable.

For example, consider a clay lens on which rests a perched aquifer. The conceptual model and therefore the hydrologic models contain this lens, but its horizontal
continuity is unknown, and it is not known whether it should be treated as a leaky layer or not. Assume that borehole ground penetrating radar is being employed to monitor the site and a recharge event occurs (Figure 2.3).

Figure 2.3 Example of one way geophysical methods can be used to calibrate hydrogeologic models.

By observing the transient evolution of water content below the clay lens, and having some knowledge of the head in the perched aquifer it is possible to determine an effective hydraulic conductivity for the clay lens.
CHAPTER 3
GROUND PENETRATING RADAR

3.1 Fundamentals of Ground Penetrating Radar

Ground penetrating radar (GPR) uses radio frequency electromagnetic waves which are sent from a source to a receiver to provide data which contain information about the material which the waves are traveling through (Bowling et al., 2005; Davis and Annan, 1989). These electromagnetic (EM) waves propagate in accordance with Maxwell’s laws (1861). When EM signals pass through conductive materials such as saline water, metal, and clay layers, the signal is attenuated and lost. These materials are classified as high loss materials (Reynolds, 1997). However, if materials are low loss (i.e. have a low electrical conductivity), the speed of these EM waves is proportional to the speed of light, and inversely proportional to the square root of the dielectric of the medium (7). In such cases, the relationship between dielectric constant and velocity can be expressed as (Davis and Annan, 1989):

\[ V = \frac{c}{\sqrt{\kappa}} \]  

(7)

Where \( c \) is the EM wave velocity through free space (the speed of light) \( (3 \times 10^8 \text{m/s}) \) and \( V \) is the velocity of the wave through a material of dielectric constant \( \kappa \).

When a traveling wave hits an interface of two or more materials that have different properties, it may be reflected, refracted, or transmitted (Reynolds, 1997)(Figure 3.1).
Direct rays represent the ray path traveling directly between the source and receiver antennae through the subsurface. Energy that travels through the subsurface to a boundary between materials with contrasting electrical properties may also be reflected to form a secondary, slower ray path between the antennae. When a neighboring material has a higher velocity than that through which the direct ray travels, refractions through the faster region result in ray-paths that are faster than direct waves.

In heterogeneous materials, $\kappa$ is an average property representing the effective behavior of the medium (Knight and Endres, 1990; Lambot et al., 2004). Generally, $\kappa$ ranges from 1 in air to 81 in water; minerals typically have dielectric constants in the range of 3-10 (Reynolds, 1997). In a mixed system, this creates a large variation in wave
speed (~0.06-0.175 m/ns) that is strongly dependant on water content (Reynolds, 1997). Traveltimes in a saturated material will be significantly longer than those in a dry soil as a result of the low velocity of EM waves in water.

Experimental efforts have developed empirical relationships between effective dielectric constants and water contents, such as (Topp et al., 1980):

$$\kappa = 3.03 + 9.3 \theta + 146.0 \theta^2 - 76.7 \theta^3$$  \hspace{1cm} (8)

$\kappa$ is the effective dielectric constant and $\theta$ is water content.

Alternatively, theoretical models, such as the complex refractive index model (CRIM) (Wharton et al., 1980), can also be used to link dielectric constant to water content.

$$\kappa_{\text{mix}}^{0.5} = \sum \theta_i (\kappa_i)^{0.5} = \phi (\kappa_w^{0.5} S + \kappa_a^{0.5} (1-S)) + \kappa_g^{0.5} (1-\phi)$$

$$= (\kappa_w^{0.5} - \kappa_a^{0.5}) \theta_w + \kappa_a^{0.5} \phi + \kappa_g^{0.5} (1-\phi)$$  \hspace{1cm} (9)

Where $\theta_i$ is the volume fraction of constituent “i” in a heterogeneous material.

For example, there are three components in a simple unsaturated material: air, water, and mineral grains. If porosity and dielectric constants are known, water content can be determined. If material volume fractions are unknown, but can be estimated, CRIM can provide an order of magnitude estimate of the water content. More generally, CRIM often provides the motivation for the use of two or three parameter empirical models for the water content-dielectric constant relationship. Whether theoretically or empirically based, the rock physics link between dielectric constant and water content is critical to making cross-borehole GPR a valuable tool for monitoring subsurface moisture.
Borehole ground penetrating radar (BGPR) transmits a high frequency radar (EM) wave through the subsurface from source antennae in a borehole to receiving antennae, which is typically in a different borehole (Reynolds, 1997). Data collected is the energy arriving at the receiver as a function of time (typically measured as a voltage). From this, traveltimes can be picked for different arrivals. Typically the direct arrival is used for traveltime tomography. Given a known ray path between the antennas, the traveltimes can be used to determine the average velocity of the propagating EM wave. The velocity of EM wave propagation is closely related to subsurface properties, particularly water content.

Standard BGPR profiles include zero offset profile (ZOP) and multiple offset gather (MOG), which is also referred to as a multi-offset profile (MOP) (Figure 3.2). In a ZOP profile, the source and receiver antenna are moved up or down the boreholes in parallel spacings. This provides a quick one-dimensional overview of the subsurface. The disadvantage of this approach is that any vertical structures that are smaller than the spacing will either not be imaged or not be laterally constrained, because the horizontal ray paths average over the area of interest.

MOG surveys emit an EM wave from the source antenna and monitor the propagation of the wave at multiple receiver locations in a separate borehole. The transmitter antenna is then moved to the next source position and the measurements are repeated (Figure 3.2). This method is more time consuming, but allows for much better coverage, and can better define small anomalies (Huismann et al., 2003). This figure uses a straight ray approximation of wave travel consistent with waves being transmitted
though the subsurface. Straight rays are used to describe a wave that is not reflected or refracted; the ray path is a line directly between the source and receiver locations.

Figure 3. 2 Schematic of typical GPR borehole arrays (Huismann et al., 2003).

3.2 Ground Penetrating Radar forward models

Ground penetrating radar datasets can be synthesized readily for application in the calibration of hydrologic models. This section will discuss the theory behind these models, create a simple GPR model, and describe its implementation.

If subsurface properties are already well constrained and a spatial map of dielectric, electrical, and magnetic properties exists, an analytical solution to Maxwell’s equations can be determined to predict the full arrival signal. Numerical models which fully represent Maxwell’s equations can also be employed to account for heterogeneity. In practice, the solution of Maxwell’s equations is not computationally economic, so
other methods are typically used. Other GPR models include using curved, or “fat” rays (Day-Lewis et al., 2005) to capture some of the impact of heterogeneity on subsurface velocity on traveltimes. Fat rays calculate the traveltime through an average area perpendicular to the wave motion, and can potentially closer approximate actual first arrival times of field data. The straight ray approximation is simpler to model and will be used in this exploratory study.

To create a straight ray traveltime model the subsurface is first divided into a series of grid blocks. The path length for each ray in a multiple offset gather array is calculated for each grid block. Each path segment is multiplied by the slowness (1/velocity) value for the grid block in which the segment is contained to determine traveltime per grid block. For each ray, these traveltimes are summed to produce the synthetic BGPR traveltime data set. The model can be described by a linear matrix equation.

\[ \hat{t} = L s \]  

(10)

where \( \hat{t} \) is a \( N_d \times 1 \) vector containing the calculated traveltime for each measurement ray path, \( L \) is a \( N_d \times N_m \) line segment matrix with element \( L_{ij} \) equal to the distance that ray \( i \) passes through cell \( j \), \( s \) is a \( N_m \times 1 \) vector in which each element, \( s_j \), represents the slowness in model cell \( j \), \( N_d \) is the total number of measurements performed in the survey, and \( N_m \) is the total number of cells in the model grid.
3.3 Geophysical Inversion

Geophysical inversion is the formal process by which the spatial distribution of slowness is estimated from the GPR traveltime measurements obtained in MOG surveys. In this way, inversion is the opposite of forward modeling discussed in the last section, which predicted GPR traveltimes given the spatial distribution of slowness. There are several important aspects to formulating an inverse problem, including selection of a grid to discretize the subsurface region of interest, establishment of quantitative criteria (i.e., an objective function) to describe data misfit and the degree to which an estimated slowness distribution agrees with prior information, and implementation of optimization techniques to find the best slowness distribution subject to criteria established in the objective function. The objective function is designed to balance minimizing data misfit \( E_d \) with honoring prior constraints on the model (Menke, 1989; Scales, 2001) \( E_m \).

\[
E(\hat{s}) = E_d + \alpha E_m
\]  

(11)

Where \( E(\hat{s}) \) is the total misfit for a candidate model representing the subsurface slowness distribution, \( E_d \) is the data misfit, \( E_m \) is model error, and \( \alpha \) represents the importance of the model error relative to data error.

\[
\begin{align*}
E &= \Sigma e_i^2 = \Sigma (d_i^o - \hat{d}_i)^2 \\
E &= e^T e = (d - \hat{d}(\hat{s}))^T (d - \hat{d}(\hat{s}))
\end{align*}
\]  

(12)

\( d \) is the observed traveltime data, and \( \hat{d} \) is the corresponding predicted data for the slowness distribution, \( \hat{s} \). The model error term \( E_m \) quantifies how closely the slowness model \( \hat{s} \) obeys a priori conceptualizations of the subsurface:
\[
E_m = e_m^T e_m = [F (\hat{s} - s_o)]^T [F (\hat{s} - s_o)] \\
E_m = (\hat{s} - s_o)^T F^T F (\hat{s} - s_o) 
\]  

(13)

where \(s_o\) is a reference slowness model, and \(F\) is a filter selected to enforce a priori constraints. This filter may enforce a model smallness constraint which minimizes the magnitudes of estimated slowness values, to a first derivative constraint which enforces spatial continuity, to any number of possible constraints, based on the conceptual model of the subsurface. Substituting equations 12 and 13 to equation 11, gives

\[
E = (d - \hat{d}(\hat{s}))^T (d - \hat{d}(\hat{s})) + \alpha (s - s_o)^T F^T F (s - s_o) 
\]

(14)

For BGPR, the data are traveltimes \(t\). For a linear forward model, the predicted data are given by \(\hat{d}(\hat{s}) = L \hat{s}\).

\[
E = (d - L \hat{s})^T (d - L \hat{s}) + \alpha (\hat{s} - s_o)^T F^T F (\hat{s} - s_o) 
\]

(15)

which simplifies to

\[
E = (d - \hat{s}_o^T L^T)(d - L \hat{s}) + \alpha (\hat{s} - s_o)^T F^T F (\hat{s} - s_o) 
\]

(16)

The minimum of the objective function occurs when the derivative with respect to the model parameters is zero.

\[
\frac{dE}{ds} = -L^T (d - L \hat{s}) + \alpha F^T F (\hat{s} - s_o) = 0 
\]

(17)

The slowness model that minimizes the objective function is found by solving Eq. 17 for \(\hat{s}\):

\[
\hat{s} = (L^T L + \alpha F^T F)^{-1} (L^T d + \alpha F^T F s_o) 
\]

(18)
The number of elements in the vector $\hat{s}$ is determined by the model grid, which should be decided based on GPR wavelength, antenna spacing, data error, prior information used in the inversion, and goals of the survey (Day-Lewis and Lane, 2004). If the grid cell size is too small, a large number of parameters must be determined from a fixed amount of data resulting in an increased reliance on prior information. If the grid cells are very large, the inversion may not be resolving structure (or heterogeneities), because local averages of property values will be estimated.

3.4 Borehole GPR Tomography Example

To illustrate BGPR tomography, consider a system with a background slowness of 1 and a box in the middle with a slowness of 10 (Figure 3.3), that is to be imaged with a survey using 20 sources and receivers (i.e. 400 rays) (Figure 3.4).
Figure 3.3 Box with slowness 10 in the center of a matrix with slowness of 1

This example uses the straight ray approximation of an EM wave path. To investigate BGPR imaging, assume that the subsurface properties are not known, and direct ray traveltimes have been properly picked.
To represent a data set collected in the field, equation 10 is used to produce a synthetic dataset of 400 traveltimes. If this was a data set collected in the field, at this point the data would typically be inverted to get some understanding of the structures and properties in the subsurface. To do this, the resolution needed for the study is necessary. If we make the grids too small, we will be attempting to determine properties of a large number of grid blocks with a small dataset. However, if the grids are too large, the inversion will not be resolving structure, merely providing an average property value for the entire area.

In this case, the area of interest is discretized using a 20x20 cell grid, resulting in a total of 400 unknown property values that must be estimated. A simple least squares
inversion with no \textit{a priori} constraint (i.e. $\alpha = 0$) results in a poor image of subsurface
since there is not enough information contained in the data to uniquely determine the
slowness value for each model grid cell. This is an underdetermined problem.

If an inversion is underdetermined, there are not enough data points to constrain a
unique parameter set, and a regularization or weighting term (the filter) is often added to
the objective function. This term can apply emphasis to certain data points, force the
resulting values to be close in value to their neighbors, force the rate of change between
neighbors to be minimized, or any number of other criteria (Scales, 2001). For example,
this can be accomplished by adding an identity matrix to the inverse problem (setting $F$
$=\mathbf{I}$):

$$\hat{s} = \left(L^T L + \alpha \mathbf{I} \right)^{-1} L^T t$$

(19)

In this case $s_0$ is zero such that estimated slownesses are forced to smaller values by the
identity matrix ($\mathbf{I}$), which enforces a smallness constraint. Figure 3.5 shows the slowness
model obtained using this constraint. There is a sharp contrast between the box and its
surroundings, so the upper and lower boundaries are well defined.
However, when rapid changes in values from grid block to grid block are not expected, a lateral continuity constraint is added in the objective function. This constraint minimizes the first derivative between congruent points (Bednar et al., 1992). Here $F$ is a matrix which approximates the first spatial derivative of $\hat{s}$ vertically and horizontally.

\[
\hat{s} = (L^T L + \beta F^T F)^{-1} L^T t
\]  

(20)

This $F$ matrix allows for minimization of the gradient between gridblocks, smoothing the data (Menke, 1989), and $\beta$ is a weight that represents the importance of the smoothing term.

Figure 3.5: Inverted GPR traveltimes with smallness weighting
This objective function provides a worse image of the subsurface (Figure 3.6), but if the target of the tomogram had a gradual change such as a plume with capillary fringe or dispersion, this would be a useful tool.

![Figure 3.6 Inverted GPR data with continuity weighting](image)

By combining the prior model constraints and setting weighting parameters to the importance of each coefficient, better resolution of subsurface features is possible in some cases

\[ \hat{s} = (L^T L + \alpha I + \beta F^T F)^{-1} L^T t \]  

(21)
Figure 3.7 shows the result of an inversion with equal weights on continuity and smallness ($\alpha=\beta$).

These inversion methods generally provide qualitative interpretations of the subsurface (Lane et al, 2005).
In complex field settings, model calibration using field data is useful for
determining what site-specific model parameter values will produce the most realistic
data estimates (Poeter, 2007). Automated calibration techniques, i.e., inverse methods,
generally work by minimizing an objective function that provides a measure of the
quality of a set of model parameters. For example, inverting geophysical data produces a
map of slownesses, which can be converted to dielectric constant, then empirically
related to water contents. These water content maps can then be compared to water
contents that are output from hydrologic models, using the objective function.

In general, the mean squared error (MSE) is often used as the objective function.
This statistic avoids canceling out of positive and negative error values:

\[
RMSE = \left[ \frac{1}{N} \sum_{i=1}^{N} (P_i - \hat{P}_i)^2 \right]^{0.5}
\]

(22)

where \(N\) is the number of data points, \(P_i\) and \(\hat{P}_i\) are the \(i^{th}\) observed and predicted datum
value. Optimization methods are used to calibrate a model by minimizing the objective
function (Menke, 1989).

While there are many methods of performing this optimization, a versatile and
efficient method is used by the program PEST: model independent parameter estimation
software (Doherty, 2004). PEST uses derivative based methods to rapidly reach the
parameters that result in the lowest value for \(E_d\). PEST uses both central derivative and
Levenberg-Marquardt methods to determine the optimal parameter values. Since most hydrologic problems are nonlinear, these methods calculate the Jacobian Matrix to create a pseudo-linear matrix, then use it to set a weight to force parameter picks to follow the linearized gradient of the objective function (Doherty, 2004).

### 4.1 Approaches to using geophysical measurements in hydrologic inverse problems

Development of effective methods for constraining hydrologic parameters with geophysical data is a current challenge. A common approach to estimating hydrologic properties using field measurements is sequential data fusion (SDF). SDF involves multiple steps (Figure 4.1). First, geophysical data are inverted to obtain a map of geophysical parameters. Next these geophysical parameters are converted to hydrologic properties using rock physics relationships. Finally, the geophysically derived hydrologic data can be used as a constraint in a traditional hydrologic calibration problem (Alumbaugh et al., 2002; Binley and Beven, 2003; Looms et al., 2008b).

![Flow chart describing sequential data fusion](image)

**Figure 4.1:** Flow chart describing sequential data fusion (Moysey et al., 2006)
Problems in estimation arise since soil properties cannot be directly measured by geophysical methods. However, empirical relationships linking geophysical and hydrologic properties do exist, for example, Topp’s relationship (equation 8) is commonly used (Topp et al., 1980). The most widely used method of estimating hydrologic properties using geophysical data involves inverting geophysical data to a map of petrophysical properties, then applying empirical formula to obtain a map of hydraulic properties (Bednar et al., 1992). This has major limitations as the inversion is underdetermined, thus a priori information used to stabilize the inversion effects the estimated properties (Moysey et al., 2006). Empirical relationships are typically developed at a laboratory scale and are often scale dependant, causing errors in the estimation process (Moysey et al., 2005). Furthermore, this method requires that inverted data that has been biased by a priori information and weighting schemes be used as a data proxy which is used in parameter estimation (i.e. fitting inverted water content values to an output of a hydrologic model).

An emerging alternative technique for constraining hydrologic properties is integrated data fusion (IDF). This approach may provide a more accurate and robust method of estimating soil properties, as a priori information is not needed to stabilize the inversion of geophysical data. A hydrologic model is coupled with a geophysical model using scale-appropriate petrophysical relationships. These coupled models are used to create a synthetic geophysical dataset for a given set of initial hydrologic model parameters. Property values in the hydrologic model (and as a result the geophysical model) are adjusted until the simulated datasets closely match the data measured in the
field survey. IDF allows information such as bore logs or estimates of soil types to be easily input into a forward model. (Figure 4.2) (Kowalsky et al., 2005).

Figure 4.2: Flow chart describing process of Integrated Data Fusion (Moysey et al., 2006)
Rucker and Ferre (Rucker and Ferre, 2004) used a method which utilizes IDF in a simulation of a zero offset profile during an infiltration event to estimate hydraulic conductivity (K), and the van Genuchten parameters $\alpha$ and $n$. The variation in dielectric constant and therefore wave velocity allowed them to constrain the wetting front location of the infiltration plume using first arrivals from refracted waves (Figure 4.3). They found that $K$ can be estimated using a bent ray approximation and GPR first arrival times, but $\alpha$ and $n$ are nonunique given only BGPR data (Figure 4.4).

Figure 4.3  Schematic representations of wetting front locations relative to zero-offset profiling (ZOP) borehole ground penetrating radar (BGPR) antennae and associated BGPR first arriving travel paths. (Rucker and Ferre, 2004)
Kowalsky et al. (2005) created a coupled geophysical and hydrologic model to estimate soil parameters for a synthetic case, then applied the models to a site at the DOE Hanford reserve. The synthetic case used a full waveform model to create a BGPR dataset while surface injection of water was simulated. These authors then used the method referred to here as IDF to estimate hydrologic parameters using a straight ray approximation of the GPR waveform (Kowalsky et al., 2005).

Looms et al (2008b) fit a ZOP GPR survey and ERT to the outputs of a 1D infiltration model. They then used a curved ray approximation of GPR traveltimes to estimate field scale hydraulic properties in an unsaturated sand during a forced solute and water infiltration event (Looms et al., 2008a). In the course of the study, they were able to refine their hydrologic model from a one layered system to a five layered model, resulting in a better fit of traveltimes (Figure 4.5).
Figure 4. The left two plots are traveltime datasets from a ZOP survey fit to a one layer infiltration model, while the right two plots are from a five layer model. From (Looms et al., 2008a)

The study found that saturated hydraulic conductivity (K) and the Van Genuchten parameter $n$ could be constrained in the upper two layers, and to a lesser extent in the third layer, but the irreducible water content($\theta_{ir}$) and $\alpha$ parameters could not be constrained.
CHAPTER 5
SYNTHETIC STUDIES OF BGPR-MONITORED INFILTRATION

This chapter lays out the synthetic study which is undertaken to explore the SDF and IDF methods of estimating hydrologic parameters using geophysical data. The synthetic study is idealized and intended to provide a feasibility assessment for lab and field-scale future work. First, the infiltration model is described. Next the SDF method of parameter estimation is set up, and results are discussed. The third section is the setup of the IDF method followed by IDF results. The results for the two methods are compared and other applications are discussed.

5.1. Setup of Reference Model used in Synthetic Studies

For each numerical experiment, infiltration events are simulated using TOUGH2 EOS-9 (Pruess et al., 1996) with a modified output file providing saturation values (Appendix B). The model consists of a 20 x 20 x 1 meter cross section of soil in the vadose zone. The initial saturation conditions are governed by gravity drainage. The saturation-pressure curve for the homogeneous soil is modeled with the van Genuchten equation (van Genuchten, 1980) and Mualem’s model (Mualem, 1976) is used to model the permeability-saturation relationship. Soil parameters used in this model were selected to be consistent with a clean sand based on the soil property database Rosetta (Schaap et al., 2001) and are given in Table 5.1. The center two meters of the upper boundary were set to be flux boundaries with a constant injection of water as the source.
of the infiltration plume. The injection rate was selected to be 0.1 L/s. The lower boundary is set as a water table. The upper and side boundaries are no flow. The resulting transient water content distribution for the reference case, obtained by multiplying saturation by porosity, is shown in Figure 5.1

<table>
<thead>
<tr>
<th>Porosity (n)</th>
<th>Irreducible water content (θᵯ)</th>
<th>Alpha (m)</th>
<th>m</th>
<th>Intrinsic Permeability (k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.375</td>
<td>0.053</td>
<td>3.5271/m</td>
<td>0.3144</td>
<td>7.44x10⁻¹² m²</td>
</tr>
</tbody>
</table>

Table 5.1 Soil parameter values used in simulations

Figure 5.1 Infiltration plume image every third recorded time First row left to right: 3100 seconds, 133 days, 417 days; second row left to right: 4.78 years, 6.34 years, 7.89 years; third row left to right: 9.44 years, 10.49 years, 13.4 years. Magnitudes are water content values.
Transient monitoring of the infiltrating plume with GPR was simulated using the Tough2 saturation results exported at irregular time intervals (i.e., every fifth time step of the simulation; Appendix A). For each GPR simulation, saturations were exported from Tough2 on a regular 1m³ grid (Figure 5.1). The soil porosity (0.375) was then used to calculate water contents, which were converted to dielectric constants using Topp’s equation (Topp et al., 1980) (8). The resulting dielectric constant distribution at each time was then used to simulate GPR traveltimes.

The GPR surveys in the synthetic experiments included both ZOP and MOG geometries, although the focus is on MOG. In both cases, a straight ray model was used, which disregards the effects of ray bending around zones of low velocity. The ZOP survey was performed with the antennas incremented by 1 meter spacings down the borehole. To investigate the impact of varying data coverage on the estimation results, several MOG surveys were performed with the spacing between subsequent transmitter or receiver positions in each borehole ranging from 1 meter, which results in the collection of 400 traveltime measurements in a survey, to 6.67 meters, resulting in the collection of only 9 traveltime measurements. To check for local minima in the objective function associated with nonuniqueness, I also vary the value at which the parameter estimation software (PEST) begins its analyses.

In Matlab, the output saturations from TOUGH2 are converted to water contents \( S \), which are then converted to dielectric constants using the Topp equation (8). GPR simulations are then performed assuming straight rays between source and receiver using a linear code developed in MATLAB. The use of straight rays assumes that either
direct ray traveltimes are picked or the waves follow a path with no “air wave” shortcuts and no refracted waves arriving before the direct waves. Note that these assumptions will likely not be met in real-world applications as the fastest arrival times would be related to paths that bend around an infiltrating front. In this work, however, the straight ray model is used to simplify the analysis and interpretation of the results. Use of straight rays therefore provides a ‘best case’ scenario for SDF in comparison to IDF.

To investigate the impact of data uncertainty several sets of simulations were performed by including data error, simulated by adding zero-mean, Gaussian random noise with a standard deviation equal to a fixed percentage of the mean traveltime:

\[
t_i = \sum_j L_0 s_j + e_i \left( \frac{1}{N} \sum_j L_0 s_j \right) = \sum_j L_0 s_j + e_i \bar{t} \quad (23)
\]

Where \( t_i \) is the simulated noisy traveltime, \( N \) is the number of traveltime measurements made in the survey, \( \bar{t} \) is the mean traveltime for the survey, and \( e_i \) is a value drawn from a \( N[0,1] \) normal distribution.

5.2 Estimation of Permeability using Sequential Data Fusion

The workflow for estimating permeability using SDF is shown in Figure 5.2. The dataset which is used as field data is simulated by running a straight ray GPR approximation through the output of the reference infiltration model (Section 5.1). Data are then inverted to obtain a slowness image, which is used to obtain a dielectric constant map. The dielectric constants are then converted to water content values using Topp’s
equation (8), i.e., the same relationship used in generating the reference data. The GPR-based estimates of water content are then used as calibration data to estimate the permeability of the field site using PEST, to optimize the TOUGH2 infiltration model.

Figure 5. 2 Flow chart describing sequential data fusion as used in this study

The geophysical inversion step in this procedure is performed using damped least squares (22). The objective function used in the inversion includes both model smallness (model norm) and continuity (first derivative) constraints. The regularization weighting parameters $\alpha$ and $\beta$ to set the relative importance of the smallness and continuity constraints relative to the data fitting constraint. Two methods for choosing these weights were investigated in this study. One method is to determine the weights that provide the lowest MSE between true water contents and water contents obtained from the GPR inversion. This method resulted in a model ‘smallness’ weight of $\alpha=1$ and a model ‘continuity’ weight of $\beta=10,000$. Unfortunately, this approach can only be used in
a synthetic setting. An alternative approach to selecting the regularization weights that is commonly used in practice involves picking weights based on a visual appraisal of how the water content map fits with the conceptual model for a specific site. This method resulted in a selection of $\alpha=1$ and $\beta=300$. The estimated slowness value in each grid block is converted back to a dielectric constant and then water content using the Topp equation (8) (Appendix B).

Estimation of permeability in the TOUGH2 infiltration model is carried out using the parameter estimation software PEST (Doherty, 2004). In the estimation procedure, permeability was the only unknown parameter; the model geometry, boundary conditions, and all other parameter values were the same as used in the simulations for the reference case. The misfit between water content values is minimized by varying intrinsic permeability. PEST was constrained such that the search space for estimated permeability values was limited to the range $1 \times 10^{-15} \text{m}^2$ to $1 \times 10^{-9} \text{m}^2$.

5.3 SDF Results

Figure 5.3 shows the results of the inversion resulting from a 400 ray survey and the regularization weights selected by optimal matching of the true and estimated water contents. The intrinsic permeability estimate that resulted from this survey was $2.75 \times 10^{-12} \text{m}^2$. 
Figure 5. 3 water contents estimated by inversion of GPR traveltime data using the MSE method of selecting filter weights.

For this and future cases, I define error as $100 \times \frac{\text{True}_k - \text{Estimated}_k}{\text{True}_k}$. The error for this survey is 62%.

Tests are also conducted to determine how data density affects the resulting permeability estimation by changing the number of source and receiver positions. Figure 5.4 illustrates that varying the number of source and receiver locations only slightly impacted the estimation results (Table 5.2).
Estimated permeability values are consistently lower than the true value, and percent errors range from 58% to 83%, with a mean error of 72% and standard deviation of 7.75%. This bias towards a lower estimate is apparent in a plot of the objective functions (Figure 5.4). Also apparent in the figure is the relatively broad bottomed nature of the objective functions for SDF estimations. Note also that the best estimate is in each case 75% of the maximum SSE. The insensitivity of the objective function near the true $k$ and relatively small reduction in error may result in difficulty finding the minima using automated optimization schemes.
<table>
<thead>
<tr>
<th># of Rays</th>
<th>k Estimate M²</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1.25E-12</td>
<td>83.11</td>
</tr>
<tr>
<td>25</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>49</td>
<td>1.25E-12</td>
<td>83.11</td>
</tr>
<tr>
<td>64</td>
<td>1.25E-12</td>
<td>83.11</td>
</tr>
<tr>
<td>81</td>
<td>1.25E-12</td>
<td>83.11</td>
</tr>
<tr>
<td>100</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>121</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>144</td>
<td>1.92E-12</td>
<td>74.05</td>
</tr>
<tr>
<td>169</td>
<td>1.92E-12</td>
<td>74.05</td>
</tr>
<tr>
<td>196</td>
<td>1.92E-12</td>
<td>74.05</td>
</tr>
<tr>
<td>225</td>
<td>1.92E-12</td>
<td>74.05</td>
</tr>
<tr>
<td>256</td>
<td>1.92E-12</td>
<td>74.05</td>
</tr>
<tr>
<td>324</td>
<td>2.79E-12</td>
<td>62.30</td>
</tr>
<tr>
<td>361</td>
<td>1.67E-12</td>
<td>77.43</td>
</tr>
<tr>
<td>400</td>
<td>2.75E-12</td>
<td>62.84</td>
</tr>
<tr>
<td>Mean</td>
<td>2.08E-12</td>
<td>71.91</td>
</tr>
</tbody>
</table>

Table 5.2 SDF Estimates of intrinsic permeability (m²) with varying ray density and using inversion method based on least squares.

Figure 5.5 shows the results of the inversion resulting from a 400 ray survey and the weights selected using the visual appraisal method. The intrinsic permeability estimate that resulted from this survey was 3.13x10⁻¹² m². The error for this survey is 57.8%.
Figure 5. 5 Inverted Infiltration plume image every third recorded time. Magnitudes are water content values. First row left to right: 3100 seconds, 133 days, 417 days; second row left to right: 4.78 years, 6.34 years, 7.89 years; third row left to right: 9.44 years, 10.49 years, 13.4 years

There is some improvement in the estimation of $k$ using the inversion method which uses weights chosen to qualitatively fit the conceptual model. The mean error is 6% lower than the previously described method (Table 5.2). However, the range in errors is the same as the error range for the sse chosen weights of the previous scheme.
<table>
<thead>
<tr>
<th># of Rays</th>
<th>k Estimate (m²)</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>25</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>49</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>64</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>81</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>100</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>121</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>144</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>169</td>
<td>1.88E-12</td>
<td>74.66</td>
</tr>
<tr>
<td>196</td>
<td>1.88E-12</td>
<td>74.66</td>
</tr>
<tr>
<td>225</td>
<td>1.88E-12</td>
<td>74.66</td>
</tr>
<tr>
<td>256</td>
<td>1.88E-12</td>
<td>74.66</td>
</tr>
<tr>
<td>324</td>
<td>9.00E-13</td>
<td>87.84</td>
</tr>
<tr>
<td>361</td>
<td>1.88E-12</td>
<td>74.66</td>
</tr>
<tr>
<td>400</td>
<td>3.13E-12</td>
<td>57.77</td>
</tr>
<tr>
<td>Mean</td>
<td>2.56E-12</td>
<td>65.41</td>
</tr>
</tbody>
</table>

SDF Estimates with varying data density (qualitative inversion)

Table 5. 3 SDF Permeability estimates resulting from inversion which qualitatively fit conceptual model.

5.4 Moment analysis

Given that the geophysically estimated water content values (Figure 5.3, 5.5) significantly underestimate the true water contents, often by a factor of 5, and that the plume shape is not reproduced in inversion, it is somewhat surprising that the SDF estimates of permeability are not worse. One hypothesis that might explain this effect is that permeability is sensitive to the shape of the infiltrating plume. In particular, if the center of mass of the plume is preserved, it may be possible to obtain an order of magnitude water content estimate even if the absolute magnitudes of the water contents may be strongly biased.
The first spatial moment in the vertical direction is calculated for each time step for both the true and inverted plumes. At early times, there is poor correlation, due to lack of rays penetrating the small plume (17-20 meters above water table). However, as the plume increases in size, it appears that the center of mass is progressing at the same rate in both cases (17-14 meters above water table). Furthermore, as time increases and more rays penetrate the plume, the inverted plume’s center of mass approaches that of the true case (14-11 m above water table)(Figure 5.6).

5.5 Estimation of Permeability using Integrated Data Fusion

In this study, integrated data fusion uses the same reference infiltration model as discussed previously (Section 5.1). Recall that to simulate the “field data,” saturations output from TOUGH2 are converted to water content by multiplying by the porosity, and
then converted to dielectric constants using the Topp equation (8). These values are converted to slownesses, and GPR traveltimes are calculated for the water content distribution at each time step in MATLAB using straight ray paths. These simulated traveltimes are the data used to calibrate the $k$ value in TOUGH2 using PEST. (Figure 5.7).

![Flow chart describing Integrated Data Fusion as used in this study](image)

**Figure 5.7  Flow chart describing Integrated Data Fusion as used in this study**

### 5.7 IDF Results

The 400 ray reference case resulted in an estimate of $7.20 \times 10^{-12}$ m$^2$, with a corresponding error of 2.7%. Tests are also conducted to determine how data density affects the resulting permeability estimation. This involved changing the number of evenly spaced source and receiver positions. The number of receivers is always equal to the number of sources. The intrinsic permeability estimate with IDF improved as data
density improved (Table 5.3) with percent error ranging from 3% to 93% with a mean of 12% and a standard deviation of 13.63%. Estimates are not consistently higher or lower than the true value.

<table>
<thead>
<tr>
<th># of Rays</th>
<th>K Estimate m²</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>5.00E-12</td>
<td>32.43</td>
</tr>
<tr>
<td>25</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>49</td>
<td>5.63E-12</td>
<td>23.87</td>
</tr>
<tr>
<td>64</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>81</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>100</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>121</td>
<td>7.34E-12</td>
<td>0.84</td>
</tr>
<tr>
<td>144</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>169</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>196</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>225</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>256</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>324</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>361</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>400</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>Mean</td>
<td>6.96E-12</td>
<td>5.97</td>
</tr>
</tbody>
</table>

Table 5.4 IDF Estimates of intrinsic permeability (m²) with varying ray density

The objective functions generally have scaled errors that span four orders of magnitude (Figure 5.8).
As with SDF objective functions, these are normalized to the max value of the sse. Note on the above figure the narrower minima compared to the SDF objective function shown in figure 5.4. Note that the lowest sse is $10^4$ times smaller than the maximum sse value.

**5.8 Comparison of SDF and IDF**

For the highest density case, i.e. when 20 source and receiver pairs are used, the objective functions for SDF and IDF are shown in Figure 5.9.
Figure 5.9 Scaled objective functions for SDF and IDF 400 ray estimates

The IDF objective function’s scaled sse spans four orders of magnitude (Figure 5.10). This indicates that traveltimes are more sensitive to perturbations in parameter values than water contents. This is a result of the poor water content estimates of the SDF approach.
Figure 5. 10 Log scale scaled objective functions for SDF and IDF 400 ray estimates

There is more variability to estimated values caused by ray density in SDF than in IDF (Figure 5.11). There was not a significant effect on the parameter estimate caused by varying the initial starting point in PEST.
The 2D SDF model was not able to produce an estimate for intrinsic permeability using a ZOP survey due to parameter insensitivity. The ZOP inversion resulted in a large zone of laterally averaged water contents. With an injection point in the upper center of the model, TOUGH2 was unable to recreate this map of water contents to the acceptance criterion of PEST.
5.9 Estimation of Permeability with unknown porosity

An additional set of experiments were performed to explore how the estimation of $k$ might be impacted when additional model parameters are unknown; in this case porosity. In these examples, the RMSE objective function was calculated by running TOUGH2 using 22,000 different combinations of intrinsic permeability and porosity values and creating an error surface. Figure 5.12 shows the “best case” optimization case, i.e. when the actual water contents for each permeability and porosity combination are taken directly as the simulation of TOUGH2. In this case, a minimum is apparent at the true values. The objective functions obtained based on radar data are shown for the SDF case, (i.e. using geophysically estimated water contents) in Figure 5.13 and for the IDF case (i.e. using the traveltimes as data) in Figure 5.14.
From the figures, it is apparent that SDF does not have a distinct minima near the true parameter values. In contrast, the IDF case provides an objective function similar to the “best case” scenario (Figure 5.12). This result suggests that IDF will be a superior, and perhaps near optimal approach to geophysical data integration.
Figure 5.14 is the RMSE of the synthetic field data fit with a traveltime dataset through the raw TOUGH2 water contents. In all three of these cases, $10^{-30}$ was added to the values so if one of the combinations of $k$ and $n$ had RMSE of zero, it would appear on this log scale plot.
5.10 Impact of Noise on permeability estimation

The final test that had significant differences was adding simulated signal noise to the traveltime dataset in the form of Gaussian error. Gaussian noise that has means that are 5%, 10%, and 15% of the mean traveltime value is also added to traveltimes to simulate nonideal conditions. Adding 5% Gaussian noise increased the SDF error by 17%, but had no affect on the IDF estimate.
Adding 10% Gaussian noise increased the SDF error by 25%, and increased the IDF error by 21% (Figure 5.15) (Table 5.5).

<table>
<thead>
<tr>
<th>% Noise</th>
<th>SDF Estimate (m²)</th>
<th>% Error</th>
<th>IDF Estimate (m)</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3.12E-12</td>
<td>57.84</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>5</td>
<td>1.88E-12</td>
<td>74.66</td>
<td>7.20E-12</td>
<td>2.70</td>
</tr>
<tr>
<td>10</td>
<td>1.28E-12</td>
<td>82.68</td>
<td>5.63E-12</td>
<td>23.99</td>
</tr>
</tbody>
</table>

Table 5.5 Results of adding Gaussian error to traveltimes

The error surface created using the SDF method does not change significantly when signal noise is added (Figure 5.16), although the magnitudes of the lowest and highest errors increases by 0.0002 and 0.0005 for 5% and 10% error respectively.
Figure 5.16 Error surfaces of RMSE water contents created using the SDF method and adding increasing signal noise

The zone of minimum values for the IDF case increased in size, but maintained its general shape. As error is increased, the minimum and the maximum values also increase (Figure 5.17).

Figure 5.17 Error surfaces of RMSE traveltimes created using the IDF method and increasing signal noise
CHAPTER SIX

CONCLUSIONS

This research set out to describe the need for and viability of a quantitative geophysical method of parameterizing and calibrating hydrologic models. Traditional hydrologic methods of determining vadose zone soil properties on the field scale are time and labor intensive, invasive, susceptible to heterogeneity, and heavily dependant on well construction or soil sampling techniques. Geophysical methods allow a large area to be studied while causing minimal disturbances to the subsurface. GPR in particular is sensitive to water content, making it an effective tool for monitoring unsaturated processes. Two methods of parameter estimation using GPR methods were studied.

Sequential data fusion consistently underpredicted the value of intrinsic permeability. Estimate results were dependant on qualitatively determined weights assigned in the inversion process, causing a 7% range in average error values. Adding small amounts of signal noise had a large effect on the estimate error, since any data collection error is amplified in the inversion process. SDF inherently includes errors which make it nearly impossible to reach a perfect fit, unless the grid blocks are so large that the information they contain becomes redundant.

Integrated data fusion is a better method for estimating intrinsic permeability in this synthetic study. The worst IDF estimate has an error 20% lower than the best SDF estimate. Low signal noise has little effect on the parameter estimate, and estimates were possible using a ZOP survey. IDF has the benefit of reaching a perfect fit if the models of the subsurface are exactly represented. While this is not a likely occurrence, the
possibility exists to precisely constrain the subsurface. Problems may arise with this method in heterogeneous studies, as the straight ray assumption becomes less valid as reflection and refraction of the GPR waves occur to a larger extent. Other limitations would be evident in an infiltration event that occurs on a timescale or physical scale that is not conducive to time-lapse GPR monitoring.

This work may not be generalizable, as it was only performed on one, simulated, homogeneous model. However, since IDF outperformed SDF in every test that was conducted on this model, it is safe to say that this study provides evidence to the benefits of IDF over SDF. To make this study completely generalizable, a large-scale stochastic project would need to have been undertaken. This project would need to involve numerous soils, heterogeneities, and geophysical techniques.

6.1 Recommendations for future work

Future work in this field should include the expansion of the method to other geophysical methodologies and full soil parameter suite estimation. Lab and field scale comparisons may address the potential limitations of the IDF method such as high loss materials, ability to account for heterogeneities, and errors due to straight ray assumptions. An additional studies on the field scale which include SDF, IDF, moment analysis as in Day Lewis (Day-Lewis et al., 2007), and traditional hydrologic methods as parameter estimation techniques would be able to conclusively determine the most efficient, effective, and accurate means of constraining parameters in the subsurface.
## Appendix A

List of recorded time steps

In seconds

<table>
<thead>
<tr>
<th>Time Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>3100.0</td>
</tr>
<tr>
<td>102300.0</td>
</tr>
<tr>
<td>3.277e+006</td>
</tr>
<tr>
<td>1.147e+007</td>
</tr>
<tr>
<td>1.966e+007</td>
</tr>
<tr>
<td>2.785e+007</td>
</tr>
<tr>
<td>3.605e+007</td>
</tr>
<tr>
<td>5.243e+007</td>
</tr>
<tr>
<td>6.881e+007</td>
</tr>
<tr>
<td>8.520e+007</td>
</tr>
<tr>
<td>1.016e+008</td>
</tr>
<tr>
<td>1.180e+008</td>
</tr>
<tr>
<td>1.344e+008</td>
</tr>
<tr>
<td>1.507e+008</td>
</tr>
<tr>
<td>1.671e+008</td>
</tr>
<tr>
<td>1.835e+008</td>
</tr>
<tr>
<td>1.999e+008</td>
</tr>
<tr>
<td>2.163e+008</td>
</tr>
<tr>
<td>2.327e+008</td>
</tr>
<tr>
<td>2.490e+008</td>
</tr>
<tr>
<td>2.654e+008</td>
</tr>
<tr>
<td>2.818e+008</td>
</tr>
<tr>
<td>2.982e+008</td>
</tr>
<tr>
<td>3.146e+008</td>
</tr>
<tr>
<td>3.310e+008</td>
</tr>
<tr>
<td>3.472e+008</td>
</tr>
<tr>
<td>3.899e+008</td>
</tr>
<tr>
<td>4.227e+008</td>
</tr>
</tbody>
</table>
Appendix B

Digital appendix
REFERENCES


67


Won, I.J., 1990. Diagnosing the Earth. Ground Water Monitoring and Remediation, 10(3): 5-&.