Assessing The Probability Of Fluid Migration Caused By Hydraulic Fracturing; And Investigating Flow And Transport In Porous Media Using MRI

James Montague

University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/graddis

Part of the Environmental Engineering Commons

Recommended Citation


https://scholarworks.uvm.edu/graddis/793

This Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
ASSESSING THE PROBABILITY OF FLUID MIGRATION CAUSED BY HYDRAULIC FRACTURING; AND INVESTIGATING FLOW AND TRANSPORT IN POROUS MEDIA USING MRI

A Dissertation Presented

by

James A. Montague

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of or Doctor of Philosophy
Specializing in Civil and Environmental Engineering

October, 2017

Defense Date: August 18, 2017

Dissertation Examination Committee:

George F. Pinder, Ph.D., Advisor
Yves Dubief, Ph.D., Chairperson
Appala Raju Badireddy, Ph.D.
Eric Hernandez, Ph.D.
Donna Rizzo, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College
ABSTRACT

Hydraulic fracturing is used to extract oil and natural gas from low permeability formations. The potential of fluids migrating from depth through adjacent wellbores and through the production wellbore was investigated using statistical modeling and predictive classifiers. The probability of a hydraulic fracturing well becoming hydraulically connected to an adjacent well in the Marcellus shale of New York was determined to be between 0.00% and 3.45% at the time of the study. This means that the chance of an induced fracture from hydraulic fracturing intersecting an existing well is highly dependent on the area of increased permeability caused by fracturing. The chance of intersecting an existing well does not mean that fluid will flow upwards; for upward migration to occur, a pathway must exist and a pressure gradient is required to drive flow, with the exception of gas flow caused by buoyancy. Predictive classifiers were employed on a dataset of wells in Alberta Canada to identify well characteristics most associated to fluid migration along the production well. The models, specifically a random forest, were able to identify pathways better than random guessing with 78% of wells in the data set identified correctly.

Magnetic resonance imaging (MRI) was used to visualize and quantify contaminant transport in a soil column using a full body scanner. T₁ quantification was used to determine the concentration of a contaminant surrogate in the form of Magnevist, an MRI contrast agent. Imaging showed a strong impact from density driven convection when the density difference between the two fluids was small (0.3%). MRI also identified a buildup of contrast agent concentration at the interface between a low permeability ground silica and higher permeability AFS 50-70 testing sand when density driven convection was eliminated.
CITATIONS

Material from this dissertation has been published in the following form:


AND

Material from this dissertation has been submitted for publication to Environmental Geoscience on September 16, 2016 in the following form:

Montague, J.A., Pinder, G.F., and Watson, T.L.. Predicting gas migration through existing oil and gas wells.
DEDICATION

To my wife, may this be but the first step in a long and happy adventure.
ACKNOWLEDGEMENTS

I would like to thank all those who have supported me and encouraged me throughout my graduate career at UVM. Thank you to the many officemates I’ve had along the way I’ve appreciated always having friendly people around to bounce ideas off of, especially Kate Johnson who has been with me through three different office moves. Thank you to the MRI research team of Richard Watts, Jay Gonyea, and Scott Hipko for your expertise and patience with MRI. Thank you to Floyd Vilmont for helping to design and build our soil columns, which so much of this work relied on. Thank you to Raju Badireddy for answering all my chemistry questions and for the use of your lab space on occasion. Thank you to my advisor, George Pinder, for giving me this opportunity to pursue research on a subject matter of my choosing and for always finding funding when we needed it most. Thank you to my committee, George Pinder, Appala Raju Badireddy, Yves Dubief, Eric Hernandez, and Donna Rizzo.

Thank you to my Parents for inspiring me to explore as a child. The annual kayaking trips fostered a love of nature and the environment. I will be forever grateful to your lessons on everything from weather patterns and clouds to how to SCUBA dive and identify fish.

A huge thank you to my wife, Marcy, for supporting me throughout this adventure and encouraging me every step of the way. Thank you for reading and proofing my work, even when the subject matter was a little dull. Thank you for giving me a supportive ear to bounce ideas off and to practice presentations on. I really appreciate everything that you have done these past 5 years.
Finally, I would like to acknowledge my funding sources. Thank you to the NSF: EAGER program for separately funding both parts of this work.
# TABLE OF CONTENTS

DEDICATION ........................................................................................................................... iii

ACKNOWLEDGEMENTS ........................................................................................................ iv

LIST OF TABLES ................................................................................................................ xii

LIST OF FIGURES ............................................................................................................... xiii

CHAPTER 1  Introduction................................................................................................................ 1

1.1  Hydraulic Fracturing ........................................................................................................... 2
    1.1.1  Well Construction ..................................................................................................... 3
    1.1.2  Targeted Formations .............................................................................................. 4

1.2  Risk of Hydraulic Fracturing ............................................................................................ 5
    1.2.1  Fluid Migration ........................................................................................................ 7
    1.2.2  Challenges to Fluid Migration ............................................................................... 7
    1.2.3  Hydraulic Fracture Extent ..................................................................................... 8
    1.2.4  Modeling Fluid Migration ...................................................................................... 10
    1.2.5  Probability Studies ............................................................................................... 11

1.3  Magnetic Resonance Imaging ........................................................................................... 11
    1.3.1  NMR Measurements in Porous Media .................................................................... 14

CHAPTER 2  Journal Paper - Potential of hydraulically induced fractures to communicate with existing wellbores ....................................................................................... 16

2.1  Abstract ............................................................................................................................. 16
CHAPTER 5  Investigation of the Buildup of Concentration at the Interface between Fine and Course Soils

5.1  Introduction ................................................................................................. 95
5.1.1  The challenges of groundwater remediation ......................................... 96
5.1.2  The effect of geologic heterogeneity on transport modeling ................. 97
5.1.3  Nuclear Magnetic Resonance in Porous Media .................................... 98

5.2  Experimental Setup .................................................................................... 101
5.2.1  Column Design .................................................................................... 102
5.2.2  Rotating the Column ........................................................................... 108
5.2.3  Salt Water Mixture ............................................................................. 109

5.3  Experimental Procedure .......................................................................... 110

5.4  Results ....................................................................................................... 110
5.4.1  Rotating Column ................................................................................ 111
5.4.2  Salt Water Column ............................................................................ 114
5.4.3  Model Results .................................................................................... 121
5.4.4  Mesh Convergence ............................................................................ 122

5.5  Discussion ................................................................................................. 123

CHAPTER 6  Conclusions .................................................................................. 125

Comprehensive Bibliography ........................................................................... 130
C.4.4  setFieldsDict ................................................................................................................. 171
C.4.5  snappyHexMeshDict ......................................................................................................... 173
C.4.6  topoSetDict ...................................................................................................................... 176
C.4.7  fvOptions ......................................................................................................................... 178
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.</td>
<td>Hydrocarbon producing formations and groups in Southwestern New York</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>(Alpha Environmental Consultants Inc., 2009)</td>
<td></td>
</tr>
<tr>
<td>4.1.</td>
<td>Model parameters. Values correspond to the Course Sand and Fine Sand.</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Highlighted values indicate that parameters are fit using optimization.</td>
<td></td>
</tr>
<tr>
<td>5.1.</td>
<td>ICP-MS results for 6 samples shown in figure 5.13. The Gd to Magnevist</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>factor is used to account for the fact that the MRI is measuring</td>
<td></td>
</tr>
<tr>
<td></td>
<td>concentration of Magnevist while ICP-MS is measuring the concentration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of only the Gadolinium which makes up 16.8% of the molecular weight of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Magnevist.</td>
<td></td>
</tr>
<tr>
<td>5.2.</td>
<td>Model parameters. Values correspond to the Course Sand and Fine Sand.</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Highlighted values indicate parameters that differ from those presented in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>table 4.1.</td>
<td></td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Total groundwater withdrawals (mgd) in 2005. Data from Maupin et al. (2014).</td>
</tr>
<tr>
<td>1.2</td>
<td>Microseismic propagation estimates for 1000 fracture stages in each of five of the most produced shale formations in the U.S. (Davies et al., 2012). Solid lines indicate the upper and lower extents of a given fracture in each respective formation.</td>
</tr>
<tr>
<td>2.1</td>
<td>Marcellus shale induced fracture heights with fit log-normal distribution.</td>
</tr>
<tr>
<td>2.2</td>
<td>Depth (ft) to shale gas formations in New York.</td>
</tr>
<tr>
<td>2.3</td>
<td>Density of existing wells located within New York, the Marcellus Shale is outlined for reference.</td>
</tr>
<tr>
<td>2.4</td>
<td>New York historical gas and oil production 1880-2014.</td>
</tr>
<tr>
<td>2.5</td>
<td>Calculation point with corresponding radius containing the nearest n wells. Hydraulic fracturing area of influence is shown in gray in the lower left of the containing circle.</td>
</tr>
<tr>
<td>2.6</td>
<td>Fit probability density function of GMM Distribution using 3 Gaussians normalized by the number of wells.</td>
</tr>
<tr>
<td>2.7</td>
<td>Average probability of model configurations holding n constant at 60 wells (top pane) and the average probability of model configurations holding $A_{hf}$ constant at 12,000 ft$^2$ (bottom pane). The three lines refer to where fractures were initiated from, the top or bottom of the target formation, and whether upward fracture growth was included.</td>
</tr>
<tr>
<td>2.8</td>
<td>Percent probability of encountering an existing well with a new horizontal hydraulic fracturing job in New York State underlain by the Marcellus Shale.</td>
</tr>
</tbody>
</table>
HF area of influence is 12,000 ft$^2$ and the number of wells in each calculation point is 20, 60, and 120 respectively. Contours are delineated at the color bar labels. ................................................................. 39

2.9. Percent probability standard deviation over a single parameter, $A_{hf}$ or n, while keeping the other constant for the Marcellus region of New York. Contours are delineated at the color bar labels. ................................................................. 40

3.1. Location of the test area in Alberta, Canada. ........................................................ 50

3.2. Number of wells missing the given property. ....................................................... 53

3.3. Confusion matrix. ............................................................................................... 57

3.4. Classification assessment for each predictive model used for training data only. Training data includes all well properties and maintains the overall distribution of failed to non-failed wells as the overall data set. Models used are: random guess (RG), back-propagation with Bayesian regularization (BG), back-propagation with gradient descent (GDX), back-propagation with Levenberg-Marquardt optimization (LM), random forest (RF), support vector machine (SVM), generalized regression neural network (GRNN), and generalized linear regression (GLR). ................................................................. 61

3.5. Comparison of SVM performance over multiple ratios between the two classes. 1 represents an equal ratio of classes, <1 indicates more wells without GM/SCVF in the training data, and >1 indicates more wells with GM/SCVF in the training data. ................................................................. 62

3.6. Classification assessment for each predictive model based on training data only. The training data is more evenly distributed between the two classes with a ratio of 1:1.25 for wells with GM/SCVF to wells without GM/SCVF. 63

3.7. Comparison of assessment metrics for validation data for models trained with and without fluid properties positive values represent metrics that are higher for models trained with fluid properties as inputs and negative values represent metrics that are higher for models trained without fluid properties. 65
3.8. Comparison of models trained with and without fluid properties for predicting wells in which all properties are observed........................................................... 67

3.9. The relative importance of each feature as described by 30 runs of a random forest model. ......................................................................................................... 68

4.1. As built schematic of the plastic sand column.................................................... 76

4.2. Pre- and post-pumping of the column after 40 days. January 28, 2015. Images are 2D slices that are 10 mm thick. Distortions around the edges appear at the edge of the field of view where the RF transmit field becomes non-uniform. ..... 80

4.3. Sagittal plane concentration scan of sand column. Gravity is from top to bottom. The concentration of Magnevist is in g/L on a log-scale. A) 0 days after flushing B) 16 days after flushing C) 26 days after flushing D) 50 days after flushing. The color bar is presented in log scale to show better depict the plume of Magnevist coming from the fine-silica cylinder........................................ 83

4.4. Sagittal plane model simulation of sand column. Gravity is from left to right. The concentration of Magnevist is in g/L on a log-scale. A) 0 days after flushing B) 16 days after flushing C) 26 days after flushing D) 50 days after flushing. The arrows represent flow direction at from the tail of the arrow. The color of each arrow represents the velocity in m/s on a log scale. ............... 85

4.5. Comparison line scans across the middle of the column for MRI results and model results. The concentrations for the model are broken into solid phase and liquid phase components. The sum of these two phases (black line) is comparable to the MRI line scan (blue circles). The concentration of Magnevist is in g/L. A) 0 days after flushing B) 16 days after flushing C) 26 days after flushing D) 50 days after flushing............................................................. 87

4.6. Resulting coefficient of variation for the modeled concentration results when each parameter is perturbed by a normal distribution defined by a coefficient of variation of 0.1. The c and f after each parameter indicate whether it is in the course or fine grained porous media.......................... 90
4.7. Coefficient of variation for the resultant concentration at each time step along the centerline of the column. The bottom of each line represents the outer edge of the course silica in the direction of gravity and the top represents the center of the fine-grained silica. Each column represents an individual parameter perturbed with a normal distribution centered at the optimized value with a coefficient of variation of 0.1. The red line shows a coefficient of variation of 0.1 for a given parameter. Each row is a different model time that corresponds to the times presented in figure 4.3................................................... 91

5.1. Theoretical breakthrough curve and actual breakthrough curve related to the volume of fluid pumped................................................................. 97

5.2. As built schematic of the plastic soil column. ......................................................... 101

5.3. Components of the soil column. ......................................................................... 102

5.4. End of the soil column. The plastic washers and alignment marker used to align the column in the same position on the MRI table for each scan. ............ 104

5.5. Frozen silica placed into the center of the column. ............................................ 106

5.6. PVC pipe filled with fine silica surrounded by saturated course sand. .......... 107

5.7. Device used to slowly rotate the soil column. ...................................................... 109

5.8. $T_1$ to concentration scaling factor found using varying concentrations of Magnevist in pure water and test tubes containing fine silica. ...................... 111

5.9. Sagittal plane concentration scan of the rotated column over a three-week period. The concentration of Magnevist is shown in g/L. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection........................................... 113

5.10. Horizontal line scan of 10 vertically averaged voxels taken across each pannel of figure 5.9. A) 0 days after initial injection of Magnevist. B) 4 days after
5.11. Sagittal plane concentration scan of the column with salt water over a three week period. The concentration of Magnevist is shown in g/L. A) 0 days after initial injection of Magnevist. B) 9 days after initial injection. C) 17 days after initial injection. D) 21 days after initial injection. ........................................ 116

5.12. Horizontal line scan of 10 vertically averaged voxels taken across each pannel of figure 5.11. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection. ......................................................... 117

5.13. ICP-MS sample locations from central cross section of the soil column from figure 5.11 and 5.12. .................................................................................................................. 118

5.14. ICP-MS concentration of Magnevist compared to MRI measured Magnevist concentrations 4 days prior. ................................................................................................. 120

5.15. Observed MRI results of the saltwater column, shown as blue circles, are compared to the modeled total concentration in black, the fluid phase concentration in magenta, and the solid phase concentration in green. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection. ... 121

5.16. Infinity norm of the modeled concentration with successive mesh refinements. The y-resolution refers to the number of volumes in the y-axis. The ratio of the y- and z- axis was kept constant with each refinement. ........... 123

6.1. Percent probability of encountering an existing well with a new horizontal hydraulic fracturing job in New York State underlain by the Marcellus Shale. There are 60 wells in each calculation point and the HF area of influence is 1,500 ft², 30,000 ft² and 1,600,000 ft², respectively. Contours are delineated at the color bar labels and are in log scale. ................................. 127

C.1. OpenFOAM development history and break points. ........................................ 162
CHAPTER 1 Introduction

Groundwater resources are crucial for agricultural production and drinking water. Resources are not replenished quickly, and once contaminated, take a lot of time and money to remediate. Understanding the risk posed to groundwater resources and having the proper tools to model contaminant movement are useful in sustaining this resource.

Groundwater withdrawals in the U.S. make up about 20% of the total freshwater withdrawals (Maupin et al., 2014). The majority of those withdrawals (68%) are used for agricultural use. Furthermore, 42% of U.S. irrigation relies on groundwater withdrawals. The reliance on groundwater in the U.S., especially in the plains and California (figure 1.1), means that significant risk to water resources should be avoided to protect food sources.

Figure 1.1. Total groundwater withdrawals (mgd) in 2005. Data from Maupin et al. (2014).
This dissertation will investigate the potential impacts posed by hydraulic fracturing on groundwater resources from deep sources and will look at a new method for investigating contaminant transport between fine-grained lenses and coarse aquifers at the lab scale using magnetic resonance imaging.

1.1 Hydraulic Fracturing

Hydraulic fracturing is a process developed in 1947 to increase the permeability of geologic formations to enhance the recovery of oil, gas, or water (Montgomery and Smith, 2010). Permeability is enhanced by injecting fluid under high pressure into the target formation. Fractures develop when the fluid pressure exceeds the minimum principal stress of the formation. After fractures have formed, the fluid pressure is reduced to allow for production from the newly enhanced formation. The drop in fluid pressure below the minimum principle stress of the formation causes the fracture walls to close (Brannon and Pearson, 2007).

Proppants, generally screened sand in the form of quartz crystals, are placed into the fluid during fracture stimulation to prevent fracture walls from closing (Howard and Fast, 1970). Proppants have since been designed for the specific reservoir rock and depth needed with many containing high amounts of Al$_2$O$_3$ or ceramics (Fitzgibbon, 1984; Lunghofer, 1985). The rate of fluid production by a fracture is determined by the hydraulic conductivity of the proppants in the fracture. The permeability of the proppant is multiplied by the propped fracture width to determine the hydraulic conductivity of a fracture (Brannon and Pearson, 2007).
Fractures are propagated in stages along the length of the wellbore. Each stage is isolated so that the pressurized fracturing fluid and proppants are directed to the appropriate location. The progress of the fracture stage is monitored by the pressure at the center of the fracture stage, referred to as the bottomhole pressure (BHP) (Martin and Valkó, 2007). The fracture propagation lengths are determined by the target formations strength, the natural fractures present, the BHP, and the volume of fluid pumped (Flewelling et al., 2013).

Fracture propagation is tracked by microseismic monitoring, which records seismic energy released by the change in stress that occurs during fracturing. An array of geophones and accelerometers is placed in a nearby well to the targeted fracture job (Warpinski, 2009). Typical fracture lengths ranged from 0 to 500 meters in 5 formations measured by Fischer and Warpinski (2012) using microseismic data from 1000 fracture stages for each formation. The measured formations were the Eagle Ford, Woodford, Barnett, Marcellus, and Niobrara Shales.

1.1.1 Well Construction

A single pad or surface location is typically used for several large scale horizontal hydraulic fracturing wells. The number of horizontal wells per pad ranges from 1 to as many as 12 (Manda et al., 2014). Multiple wells emanating from the same pad means that the casing must be deviated from vertical to provide enough space between the horizontal sections of each well to maximize the producible area while preventing wells from competing with each other due to intersecting fracture regions (Yu and Sepehrnoori, 2013).
Wells are drilled using multiple casings, set from largest to smallest, as the well is drilled deeper. A surface casing is installed to below the deepest formation with fresh groundwater and then the annular space, the space between the well casing and the outside borehole, is filled with cement (U.S. Department of Energy, 2009). Cement is used to isolate formations from each other and the well casing with the goal of preventing fluid migration between formations. After the cement has set around the surface casing, further well casings are placed and cemented, ultimately terminating with the production casing, which is set into the target formation, either vertically or horizontally. The production casing is perforated using a perforating gun with holes spaced at specified angles to enhance the maximum volume of induced fractures (Abass et al., 1994).

1.1.2 Targeted Formations

Hydraulic fracturing is used to enhance the conductivity of low permeability formations, most commonly shales that contain a high organic content, which allow for thermogenic methane to form. The most productive shale formations in the U.S. are the Marcellus Shale, the Haynesville Shale, and the Barnett Shale (U.S. EIA, 2016). These formations are targeted by drilling vertically until a desired depth, the kickoff depth, is reached. Once the kickoff depth is reached, the well is directionally drilled into the formation. Other formations targeted for hydraulic fracturing include tight gas formations and coal bed methane. Tight gas refers to sedimentary rocks with a permeability lower than 0.1 mD that are not shales (Law and Curtis, 2002). Coal bed methane refers to gas that is produced by lowering the pressure within a coal formation to allow desorption of
natural gas from the rock matrix. Typically, this involves dewatering the cleats (natural fractures) (Schein and Mack, 2007).

1.2 Risk of Hydraulic Fracturing

The topic of hydraulic fracturing risk has been a contentious issue since the early 2000’s when natural gas production by hydraulically fractured wells began to control a large share of the natural gas and oil markets in the U.S. (U.S. EIA, 2016). In December of 2016, the U.S. EPA issued a final report on the impact of hydraulic fracturing on drinking water resources (U.S. EPA, 2016). This report covered the entire life cycle of a hydraulically fractured well from the acquisition of water to the disposal and storage of produced water coming back from the well. They concluded that there is some risk from the hydraulic fracturing water cycle, especially in regards to improper disposal, treatment, and storage of waste fluids resulting in potential contamination of surface and groundwater near hydraulic fracturing operations.

This dissertation focuses on the risk of injection of fracturing fluid and the integrity of wellbores. These represent two of the three identified flow paths from injecting fracturing fluid. The pathways identified in the EPA report include: 1. Through the rock and soil matrix (including natural fractures, faults, and bedding planes); 2. Through a compromised production well casing, either within the annulus or between the casing and the surrounding soil and rock; 3. Through existing wellbores close to the production well (Harrison, 1983; 1985). Paths 2 and 3 require that the well casings or plugs be compromised for flow to be possible (Dusseault and Jackson, 2014), while path 1 is highly dependent on regional geology and requires some degree of localized discontinuities within
the layers (Leff, 2011; Flewelling et al., 2013). Individual well casings can be tested for vent flow or pressure build up at the well head, which is indicative of failure (Watson and Bachu, 2009). These flow paths can work independently or in concert to form a pathway to the surface, provided an upward pressure gradient exists.

Paths 1 and 3 require a discontinuity to exist between the targeted formation and shallower groundwater resources. Therefore, the distance between targeted formations and groundwater resources is a contributing factor in whether fluid is able to migrate upwards (Reagan et al., 2015; Jackson et al., 2013b). The typical distance between targeted shale formations and groundwater resources is on the order of 1,000 feet, but can be as small as 100 feet in some formations (U.S. EPA, 2016). Larger distances reduce the likelihood of induced fractures connecting the target formation to groundwater resources due to fracture growth being limited by overlying formations (Fisher and Warpinski, 2012). Microseismic data on fracture height growth in various shales in the U.S. has identified fracture growth that exceeds 1,000 ft.; however, the number of fractures exceeding 1,148 ft. is only 1% (Davies et al., 2012). In the absence of naturally occurring discontinuities, there is a chance that a nearby existing well may intersect induced fractures. This chance is further enhanced in areas where there has historically been oil and gas extraction in the target formation or those formations near the target formation. This is further investigated in Chapter 2 for the Marcellus Shale in New York.

The second pathway along or through the casing of the production well is caused by construction defects, or cement and casing degradation over time. Typically, multiple casings are constructed and the annulus is filled with cement. This is done to isolate vari-
ous formations including water bearing formations and the target formation (*Ross and King*, 2007).

### 1.2.1 Fluid Migration

Fluid migration, primarily in the form of natural gas, has been investigated in areas where hydraulic fracturing has taken place. Identifying the source of natural gas can be done by examining the concentrations and isotopic composition of the hydrocarbons, specifically the ratio of stable carbon isotopes, known as $\delta^{13}C$, and the ratio of methane to higher-chain hydrocarbons (*Rowe and Muehlenbachs*, 1999; *Schoell*, 1980). Sourcing natural gas can demonstrate that a pathway exists between the source formation and the formation that the gas was found in, but it does not illuminate the location or type of pathway that is present.

The presence and source of natural gas in groundwater has been investigated in the Marcellus Shale regions of Pennsylvania. Researchers have found that the presence of thermogenic methane, generally sourced from deeper formations, is more likely to appear in groundwater that is close to hydraulic fracturing operations (*Darrah et al.*, 2014; *Jackson et al.*, 2013a; *Osborn et al.*, 2011). *Hammond* (2016) extended this to determine that gas in water wells near Dimmock, PA was thermogenic, but from shallower source formations than the targeted Marcellus Shale.

### 1.2.2 Challenges to Fluid Migration

Vertical migration of fracing fluid or brine requires an upward pressure gradient to overcome gravitational forces and capillary forces. *Flewelling and Sharma* (2014)
place upper bounds on the upward gradient that is dependent on the maximum pore pressure that overlying cap rock formations can sustain before fracturing and releasing the built-up pressure. This also means that sustained pressures are generally low in areas that do not naturally experience upward fluid migration prior to fracking. They further note that gas production will create a pressure gradient in the direction of the producing well and hinder upward fluid migration. Natural gas, on the other hand, does not require a positive pressure gradient to move upwards because it is naturally buoyant.

Prior to fracking, gas is trapped within isolated pores of the target formation. The fracturing process increases the permeability of the target formation and mobilizes trapped gas to flow primarily towards the producing well (Bažant et al., 2014). Some gas is not captured by producing wells and is available to migrate upwards; however, barriers in the form of overlying low permeability formations and capillary forces, hinder migration through the overlying soil and rock matrix.

1.2.3 Hydraulic Fracture Extent

The upward propagation of hydraulic fractures is typically measured through microseismic monitoring (Maxwell, 2011). Microseismic monitoring detects shear waves that are generated during the tensile opening of fractures when fluid pressure exceeds the minimum principle stress of the rock. In this case, the fracture is dilated normal to the fracture surface. Shear waves can also be generated by the shear slipping in which the movement is parallel to the fractures. Finally, shear waves can be generated by the contraction of the rock or fracture where the fracture aperture is decreased by surrounding pressure (Busetti et al., 2014). Perturbations in the stress field of the surrounding rock can
be caused by poroelastic coupling of tensile and shear fractures causing microseisms to be detected beyond the fractured region (Warpinski, 2009). Microseisms provide an estimate of the extent of fracture propagation and volume of fractures due to hydraulic fracturing, but may overestimate the maximum growth of fractures (Warpinski et al., 2004). Fisher and Warpinski (2012) used microseismic data from individual fracture stages to estimate the length that fractures propagated in various shale formations around the U.S., including the Marcellus Shale. Figure 1.2 shows the distances that they estimate, ranging from a few meters to around 500 meters, with most fractures propagating between 50 and 100 meters.

Figure 1.2. Microseismic propagation estimates for 1000 fracture stages in each of five of the most produced shale formations in the U.S. (Davies et al., 2012). Solid lines indicate the upper and lower extents of a given fracture in each respective formation.
1.2.4 Modeling Fluid Migration

Numerical modeling of hydraulic fracturing has used possible worst-case scenarios to develop upper bounds on fluid volumes capable of migrating into shallow aquifers. (Lange et al., 2013; Kissinger et al., 2013). Worst-case scenarios qualitatively provide information on the extent of fluid movements based on possible, but unlikely geologic scenarios; however, they do not provide the quantitative measures of the probability of occurrence. Modeling is helpful in analyzing the conditions required for migration to be possible. A second modeling attempt used Modflow to model upward flow enabled by fractures that propagate out of the target formation and increase the conductivity of overlying formations (Myers, 2012). This study, however, has been highly criticized for using unrealistic boundary conditions for both pressure and flow within the shale (Cohen et al., 2013).

Finally, modeling was conducted by Reagan et al. (2015) to examine the short-term transport of fluids from a single hydraulic fracturing well through various pathways. Their pathways included fractures extending to the surface or intersecting natural fractures as well as fractures intersecting existing wells or the fracturing well. The study makes it clear that the probability of each scenario was not addressed, merely the resulting quantity of fluid delivered to the shallow water resource. They found that W-cases transport larger amounts of methane to aquifers than F-cases. In addition, both cases are highly dependent on the production regime within the tight gas reservoir (producing wells limit the amount of gas available for transport). Furthermore, the amount of gas available for migration is limited by slow recharge from the source rock, meaning that the amount
initially available for migration is constrained to the volume of gas stored within the frac-
tured portion of the reservoir.

Modeling by Lange et al. (2013), Myers (2012), and Reagan et al. (2015) have
shown that modeling fracturing fluid migration requires many assumptions for geologic
conditions, formation properties, and boundary conditions. For these reasons, quantitative
measures of risk are difficult to extract without conducting extensive geologic surveys.

1.2.5 Probability Studies

A study to determine the risk of water pollution associated with natural gas ex-
traction utilized probability bounds analysis and probability boxes in an attempt to quan-
tify the risk fracturing poses to water resources (Rozell and Reaven, 2012). The study ana-
alyzed multiple aspects of the risk, from truck crashes and spills to fluid flowing upward
through the rock matrix. For several aspects of the study, the probability boxes had tight
bounds associated with low degrees of uncertainty; but for others, such as the probability
of fracturing fluids flowing though the rock matrix, the probability bounds were quite large
indicating very high uncertainty. Their study does justify further study to reduce the un-
certainty of their pathways.

1.3 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is based on nuclear magnetic resonance
(NMR), which uses the process of absorption and reemission of radio waves by the nuclei
of atoms in the presence of a magnetic field. MRI relies on the spin or angular momentum of an atom’s nuclei, in particular the spin of hydrogen protons. A hydrogen atom’s
charge causes it to behave as a microscopic bar magnet (Hall, 1964). This means that hy-
drogen atoms placed in a strong magnetic field quickly align with the magnetic field lines. The time required for this alignment to occur is referred to as $T_1$. The measurement of $T_1$ is done by pushing the hydrogen atoms out of alignment and measuring how long it takes for them to return to equilibrium. This time is affected by the environment where the hydrogen atom is located, which is how different tissues in the human body can be detected using MRI.

MRI images are created by measuring the magnetic properties of hydrogen atoms at varying locations. Magnetic gradients are used to linearly vary the Larmor frequency, or frequency that an atom precesses at in a magnetic field (Mansfield and Maudsley, 1977). Precessing is the slow movement of a spinning body around another axis caused by the action of another force, for example a top that is rotating around an axis begins to precess as gravity acts on it. This causes the top to wobble slowly. Three steps are needed to get 3D spatial information. First, the Larmor frequency is varied linearly along the $Z$-axis by generating a magnetic gradient in that direction. A slice is selected by generating the appropriate resonant radio frequency (RF) pulse for the magnetic field strength at the desired $Z$ location. Once a slice has been excited, a second magnetic gradient is generated along the $Y$-axis. This encodes the phase along rows within the selected slice by dephas- ing the spins along the $Y$-axis. Finally, a gradient is generated along the $X$-axis while the RF receiver coil records the signal that the hydrogen atoms reemit. The measured signal will contain a range of frequencies linearly dependent on their location within the final gradient in the $X$-direction and the Larmor equation. The signal coming from areas with lower magnetic field strength will have a lower frequency than the signal coming from areas with higher magnetic field strength. The exact location of the signal can be deter-
mined from the Fourier transformation of the signal to isolate specific frequencies and therefore X-locations. By taking multiple observations of the phase encoding, the contribution of each phase to the signal of a given frequency allows the Y-location to be determined (Edelstein et al., 1980).

Resonant RF pulses cause the magnetization of the hydrogen atoms to flip by a given angle (the flip angle), and to precess like a gyroscope that has been bumped off its primary rotational axis. Once the radio waves are no longer being pulsed, the hydrogen atom will continue to precess and induce a measurable signal in the receiver coil that decays over time. This signal is referred to as $T_2$ decay.

The $T_1$ and $T_2$ signals are caused by separate processes. The $T_1$ signal is the recovery of the hydrogen magnetization back into equilibrium and the $T_2$ signal is caused by the decay of the hydrogen atoms spin and the dephasing of individual atoms caused by slight differences in the magnetic environment. $T_1$-weighted images are generated by rapidly generating RF pulses and measuring the signal generated. The time between RF pulses is the repetition time (TR). Using a TR that is smaller than the time required for the atom to recover to its equilibrium provides more signal strength in materials that recover more quickly, referred to as having a short $T_1$ value.

For the work presented in this dissertation, $T_1$-weighted images are used in conjunction with a paramagnetic contrast agent which significantly lowers the $T_1$ value in areas where it is present. The $T_1$ value of the areas with contrast is linearly proportional to the concentration of the contrast agent present once the background $T_1$ value of the water and media is removed.
1.3.1 NMR Measurements in Porous Media

Nuclear magnetic resonance (NMR) has been used for non-invasive imaging of the structure of porous media. Issa and Mansfield (1994) used $T_1$ weighted images to estimate permeability in sandstones using the relationship between the $T_1$ of porous media and the $T_1$ of water. Komlosh et al. (2011) determined the average porosity of glass capillary tubes on the voxel scale (three-dimensional volume over which measurements are averaged) using double pulsed-field gradient diffusion weighted MRI. Saturation curves for various porous materials as was calculated by Muir et al. (2014) using $T_2$ mapping Spin Echo Single Point Imaging as water was displaced by heavy water.

NMR has also been used to measure flow through porous media in small-scale experiments. Shattuck et al. (1995) measured thermal convection patterns using a spin echo technique in a small cylinder packed with monodisperse plastic beads. Kimmich et al. (2001) used NMR to map fluid velocities in percolation clusters created with a circuit board plotter. Khrapitchev et al. (2002) tracked the locations of particles in a small porous media column using velocity exchange spectroscopy to examine the dispersive flow within a column. This NMR method allowed them to track molecules by applying a spin and observing their displacement after a small delay in a 10 mm diameter column. Seymour and Callaghan (1999) provide a good overview of the NMR methods used to image fluid flow in porous media.

In addition to mapping physical properties of porous media, the concentrations of magnetic tracers such as Gd-diethylenetriaminepentaacetate (Gd-DTPA$^{2-}$) can be tracked through time. Haber-Pohlmeier et al. (2010) demonstrated that Gd-DTPA$^{2-}$ can be
tracked within a sand column in an MRI scanner to determine flow and transport behaviors in three-dimensions.

Contrast agents are required to be compounds that change the relaxivity of the solution in which they are dissolved. Alternative contrast agents include NiCl$_2$ (Pearl et al., 1993) and CuSO$_4$ (Greiner et al., 1997). Greiner et al. (1997) packed two small columns with glass beads and calculated the concentration of copper sulfate at the outlet with both MRI and atomic absorption spectroscopy. The results for both homogeneous and heterogeneous bead packing showed that MRI could accurately measure outflow breakthrough curves with the added benefit of quantifying the concentration within the column.
CHAPTER 2 Journal Paper - Potential of hydraulically induced fractures to communicate with existing wellbores

James A. Montague and George F. Pinder, Civil and Environmental Engineering, University of Vermont, Burlington, Vermont, USA.

Corresponding author: J. A. Montague, Civil and Environmental Engineering, University of Vermont, 33 Colchester Ave, Burlington, VT 05405, USA (james.montague@uvm.edu)

2.1 Abstract

The probability that new hydraulically fractured wells drilled within the area of New York underlain by the Marcellus Shale will intersect an existing wellbore is calculated using a statistical model, which incorporates: the depth of a new fracturing well, the vertical growth of induced fractures, and the depths and locations of existing nearby wells. The model first calculates the probability of encountering an existing well in plan view and combines this with the probability of an existing well being at sufficient depth to intersect the fractured region. Average probability estimates for the entire region of New York underlain by the Marcellus Shale range from 0.00% to 3.45% based upon the input parameters. The largest contributing parameter on the probability value calculated is the nearby density of wells; meaning that due diligence by oil and gas companies during construction in identifying all nearby wells will have the greatest effect in reducing the probability of interwellbore communication.
2.2 Introduction

A large part of Southern New York State is underlain by the Marcellus Shale. The depth varies spatially, ranging from 3000 to 6000 feet below the surface, where gas is readily present. The large distance through overlying rock formations to shallow aquifers makes fluid migration unlikely without the presence of discontinuities to provide pathways through the rock (Leff, 2011; Flewelling et al., 2013). Unplugged existing wells or existing wells with compromised casings/cement as well as natural fractures and compromised producing wells all provide potential pathways from depth (Harrison, 1983; Nordbotten, et al., 2004; Nordbotten, et al., 2005; Dusseault and Jackson, 2014). Existing boreholes both producing and non-producing that are either unplugged or have had their casings compromised in some way provide the most likely pathway for fluid migration by providing a continuous conduit from depth (Harrison, 1985). Upward propagation through natural fractures, on the other hand, is deemed impossible by Engelder et al. (2014). The presence of large capillary forces and low water saturations that exist in the Marcellus Shale serve to trap fracturing fluid and brine within the shale matrix. For the purposes of this study we consider only the potential of pre-existing wells in combination with the induced fracture network to permit migration.

Migration in this case can refer to the migration of any fluid (gas, brine, or fracturing fluid). While it may be possible for liquids to migrate upwards, the conditions required to reach shallow aquifers or the surface seem unlikely to occur. Flewelling and Sharma (2014) provide a good analysis of the conditions necessary to mobilize fracturing fluid or brine. Gases are therefore the most likely fluid to migrate because gas requires much less pressure to mobilize due to its buoyancy.
Modelling was conducted by Reagan et al. (2015) to examine the short-term transport of fluids from a single hydraulic fracturing well through various pathways. Their pathways included fractures extending to the surface or intersecting natural fractures (F-cases) as well as fractures intersecting existing wells or the fracturing well (W-cases). The study makes it clear that the probability of each scenario was not addressed, merely the resulting quantity of fluid delivered to the shallow water resource. They found that W-cases transport larger amounts of methane to aquifers than F-cases. In addition, both cases are highly dependent on the production regime within the tight gas reservoir (producing wells limit the amount of gas available for transport). Furthermore, the amount of gas available for migration is limited by slow recharge from the source rock, meaning that the amount initially available for migration is constrained to the volume of gas stored within the fractured portion of the reservoir.

Data on New York State wells and microseismic data collected from Marcellus fracturing jobs within Ohio, Pennsylvania, and West Virginia are used herein to develop a statistical model to better understand the risk of communication from hydraulic fracturing associated with all existing boreholes and wells reported in the New York Department of Environmental Conservation (NYDEC) downloadable well database. Since the integrity of the wells and borings are not considered in this paper, the results found do not represent the actual risk associated with hydraulic fracturing by means of communication with existing wellbores, but instead place an upper bound on that estimation.
2.2.1 Fracture Propagation

Induced upward fracture growth for hydraulically fractured wells in the Marcellus region was estimated from microseismic data collected for over 1000 fracture stages by Fisher and Warpinski (2012) and digitized by Davies et. al (2012). Microseismic monitoring is commonly used during hydraulic fracturing to estimate the extent of fracture growth by detecting shear waves produced during the fracturing process. Shear waves are generated during the tensile opening of fractures when fluid pressure exceeds the minimum principle stress of the rock, in this case the fracture is dilated normal to the fracture surface. Shear waves are also generated by shear slipping in which the motion is parallel to the fractures. Finally, shear waves can be generated by the contraction of the rock or fracture (Busetti et al., 2014). Perturbations in the stress field of the surrounding rock can be caused by poroelastic coupling of tensile and shear fractures causing microseisms to be detected beyond the fractured region (Warpinski, 2009). Microseisms provide an estimate of the extent of fracture propagation and volume of fractures due to hydraulic fracturing, but may overestimate the maximum growth of fractures (Warpinski et al., 2004). This data set, therefore, provides a maximum estimate for fracture propagation distances that we use to create a probability distribution of the maximum vertical extent of fractures based on a log transformation of the propagation heights in feet and a fit log-normal distribution. The parameters that were extracted from this data provide fitting parameters of $\mu = 5.979 \ln(\text{ft.})$ and $\sigma = 0.579$ with a Kolmogorov-Smirnov goodness of fit resulting in a p value of 0.014 (figure 2.1). The parameters $\mu$ and $\sigma$ are fitting parameters defined by equation 2.1, where $m$ is the mean of the untransformed data and $\nu$ is the variance of the untransformed data.
\[ \mu = \ln \left( \frac{m}{\sqrt{1 + \frac{\nu}{m^2}}} \right), \quad \sigma = \sqrt{\ln \left( 1 + \frac{\nu}{m^2} \right)} \]  

(2.1)

The envelope of induced fractures provides a region of higher permeability in which fluid movement is possible.

Figure 2.1. Marcellus shale induced fracture heights with fit log-normal distribution.

2.2.2 Energy Production in New York State

Energy extraction has been occurring in New York since the mid 1800s. The Marcellus in particular has been produced since 1880, the depth of the bottom of the Marcellus shale are shown in figure 2.2 (Leff, 2011). However, the discovery of easily
accessible conventional reservoirs within New York led to a decrease in Marcellus production wells until recent directional drilling and slick water technologies made unconventional reservoirs more accessible. The locations, uses, dates of installation, and depths of energy wells have been recorded by New York since 1900. The density of wells across New York is shown in figure 2.3, with darker colors representing higher densities. The outline of the Marcellus Shale’s producible area is outlined for reference. Many of the wells included in the NYDEC downloadable well database targeted conventional oil and gas producing formations other than the Marcellus. Table 2.1 shows formations located in Southern New York that have been targeted for energy extraction at some point since the mid 1800s. The last column shows the relative depth with respect to the Marcellus Shale, which is the bottom unit of the Hamilton Group in this region.

![Figure 2.2. Depth (ft.) to shale gas formations in New York.](image)

---

21
The state of New York estimates that around 70,000 wells have been drilled since the 1800s; of those, a database of 41,000 wells exists. Of these 41,000 wells, 29,000 are located in the same region that is underlain by the Marcellus Shale, with well depths ranging from a few hundred feet to upwards of 10,000 feet. The number of reported wells increased significantly in the late 60s from a rate of about 100 wells per year to upwards of 800 per year, while the production of oil and natural gas dropped, as shown by the historical production rates in figure 2.4. This suggests that many of the unaccounted for wells were constructed before the 1960s. The large number of unaccounted for wells necessitates a probabilistic approach to dealing with the locations and depths of existing wells.
Table 2.1. Hydrocarbon producing formations and groups in Southwestern New York (Alpha Environmental Consultants Inc., 2009)

<table>
<thead>
<tr>
<th>Formation</th>
<th>Type</th>
<th>Production</th>
<th>Above/Below Marcellus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadaway Group</td>
<td>Sandstone and Shale</td>
<td>Gas / Oil</td>
<td>Above</td>
</tr>
<tr>
<td>Hamilton Group</td>
<td>Sandstone and Shale</td>
<td>Gas</td>
<td>At/Above</td>
</tr>
<tr>
<td>Onondaga</td>
<td>Limestone</td>
<td>Gas / Oil</td>
<td>Below</td>
</tr>
<tr>
<td>Helderberg Group</td>
<td>Limestone</td>
<td>Gas</td>
<td>Below</td>
</tr>
<tr>
<td>Clinton Group</td>
<td>Shale</td>
<td>Gas</td>
<td>Below</td>
</tr>
<tr>
<td>Medina Group</td>
<td>Sandstone</td>
<td>Gas</td>
<td>Below</td>
</tr>
<tr>
<td>Queenston</td>
<td>Sandstone and Shale</td>
<td>Gas</td>
<td>Below</td>
</tr>
<tr>
<td>Trenton -Black River</td>
<td>Limestone</td>
<td>Gas</td>
<td>Below</td>
</tr>
</tbody>
</table>

Figure 2.4. New York historical gas and oil production 1880-2014.
2.3 Methods

The large number of unaccounted for wells within New York State means that determining exactly where new hydraulic fracturing (HF) wells will intersect existing wells is not possible and necessitates a probabilistic approach to existing well locations and depths. This requires several assumptions to be made with regards to existing wells and fracture propagation:

1. The existing wells in our database are a representative subset of all existing wells, known and unknown.

2. The depth of existing wells that are close to one another come from the same distributions, i.e. wells in an area tend to target the same formation or set of formations.

3. Existing wells within the data set are vertical, the inclusion of horizontal wells within the data set would provide a greater opportunity for intersection with hydraulic fractures.

4. Fractures growth is not impacted by nearby wells or existing fractures, i.e. fractures will not bend towards nearby discontinuities to increase the volume of the fractured region.

The first two assumptions mean that to calculate the probability in a spatial sense the region must be subdivided so that distributions of well depths can be determined on an appropriate scale. The third assumption allows the probability calculation to be simplified to point intersections rather than point and line intersections. Finally, the fourth assumption eliminates the need for fracture growth modelling from the statistical model.
The probability calculation is evaluated for each subdivided region or representative area (RA). For the purposes of this analysis, RAs are located in a uniform pattern consisting of one RA centered every mile in both the x and y directions. One mile is used to provide significant overlap of each RA so that probability values transition smoothly across the region. A more efficient way to discretize may be adaptive spacing based on the size of neighboring RAs. This has the benefit of reducing the number of RAs needed for the analysis, but the calculated probabilities do not transition as smoothly across RAs. For the purposes of this analysis using more RAs did not slow down the model enough to warrant an adaptive spacing approach. However, analysis over larger regions would benefit from either adaptive spacing or from a coarser uniform discretization.

Representative areas are used to determine the probability of induced hydraulic fractures encountering an existing well. Assigned to each RA are the $n$ existing wells from the NYDEC downloadable well database that are closest to the center of the RA. We will consider these $n$ wells as being within the RA. For the purposes of this study, we investigated cases in which $n$ equals 20 wells up to $n$ equals 120 wells. In addition to knowing the wells within each RA, the radius necessary to contain these $n$ wells, $r_{RA}$, is needed to determine the density of wells in the region. Equation 2.2 defines $r_{RA}$,

$$r_{RA} = \frac{d_n + d_{n+1}}{2}$$

(2.2)

where $d_n$ is the distance from the center of the RA to existing well $n$ and $d_{n+1}$ is the distance to existing well $n+1$. This formulation of the radius of contained wells is chosen so that included wells are not on the boundary of the RA and the area is not biased towards existing well $n$ or existing well $n+1$. 

25
To simplify the probability calculation further, the probability will be determined in two separate parts and then those parts will be multiplied to get the full three-dimensional probability of a new HF well encountering existing wells. The first part is calculating the horizontal probability of encounter, or the probability of an existing well being within the plan view footprint of an HF well. The second part is calculating the probability that wells within the RA are at sufficient depth to intersect the fractured region around the HF well. The total probability will then be calculated by multiplying the horizontal and vertical probabilities.

### 2.3.1 Horizontal probability of encounter

The probability that a new HF well placed randomly within the RA horizontally intersects an existing well is dependent on the area of the RA ($A_{RA}$), the number of existing wells in the RA ($n$), and the anticipated x-y area of influence of the HF wells ($A_{hf}$). In this case $A_{hf}$ is defined as the x-y area that is underlain with induced hydraulic fractures. Figure 2.5 shows an example representative area containing $n$ wells and an HF well. The x-y HF area of influence is the gray area in the lower left quadrant of the RA, but it can be located anywhere within the circle and oriented in any direction since the probability of encountering the fractured region is calculated for the entire RA. Typically new HF wells are located perpendicular to natural fractures or parallel to the minimum principle stress in the x-y plane (Engelder et al., 2009). However, due to the large number of unidentified existing wells, the location of existing wells within an RA is random as well. This means that the orientation of the HF well does not enter into the probability equa-
tion, only its area (see equations 2.3 and 2.4). Note that this is for new, yet to be located HF wells in the region of interest.

Figure 2.5. Calculation point with corresponding radius containing the nearest n wells. Hydraulic fracturing area of influence is shown in gray in the lower left of the containing circle.

The probability of a single existing well intersecting $A_{hf}$ is simply a ratio of the two areas, $A_{hf}$ and $A_{RA}$, given as $P_{h1}$ in equation 2.3.

$$P_{h1} = \frac{A_{hf}}{A_{RA}} \quad (2.3)$$

More than one well exists in the containing circle, defined by $n$, so the probability of encountering at least one well is given in equation 2.4 as $P_{h}$.

$$P_{h} = 1 - \left[(1 - P_{h1})^n\right] \quad (2.4)$$

The bracketed terms represent the probability of the hydraulically fractured region intersecting none of the $n$ existing wells in the RA. The probability of one or more wells intersecting the fractured region is therefore the converse of the bracketed terms. The next
step is to determine the probability of the existing wells vertically encountering the fractured region.

### 2.3.2 Vertical probability of encounter

Vertical intersection of an existing well with the fractured region is determined by combining the depth to the horizontal segment of the HF well with probability distributions for regional well depths and upward fracture growth. The depth to the horizontal segment of the HF well is taken as a single value based on the depth of the Marcellus Shale at the center of the RA. Probability distributions are fit to the depth of wells for the \( n \) wells within each RA, while the probability of fracture height is represented by the fit lognormal distribution for the whole of the Marcellus shale.

A Gaussian mixture model (GMM), given by equation 2.5, is used to fit depth data, where \( \mu_j \) are the means associated with each Gaussian, \( \sigma_j^2 \) are the variances, and \( \pi_j \) are the weights (which sum to one). A GMM allows for multi-peaked data to be fit. This makes it ideal for fitting the depth of existing wells that have targeted multiple formations in the same regions over the history of oil and gas drilling in New York. An example distribution is shown in figure 2.6, in which 3 Gaussians are fit to 60 well depths.

\[
p(y | \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2, \pi_1, \ldots, \pi_k) = \sum_{j=1}^{k} \pi_j N(\mu_j, \sigma_j^2) \quad (2.5)
\]

GMM fitting is done using the EM algorithm which uses maximum likelihood estimates of the GMM parameters given the desired number of Gaussians, \( k \) (McLachlan and Peel, 2000; Dempster et al., 1977). The Matlab function gmdistribution.fit is used to accomplish this task for values of \( k \) ranging from 1 to 20. The value of \( k \) to be used is calculated
with the Akaike information criterion (AIC). The AIC identifies the simplest model that fits the data presented by rewarding the maximum likelihood calculated for a given number of independent parameters, k, while penalizing large values of k used in the GMM (Akaike, 1974). Using the GMM with a k that produces the lowest AIC ensures that the best fit for the data is used without overfitting.

Figure 2.6. Fit probability density function of GMM Distribution using 3 Gaussians normalized by the number of wells.

The probability that an existing well will be at or below the fractured region is calculated first with the convolution of the fracture propagation distribution and the well depth distribution and then a cross-correlation of the horizontal segment depth of the HF well and the result of the convolution integral (Casella and Berger, 2002). Equation 2.6 provides G which is the random variable for the distance between the base of the well and the top of the fractured zone given the random variables for the depth of the horizontal portion of the fracturing well, S, the fracture propagation height, F, and the depth of ex-
isting wells, $D$. The probability of the fractured region vertically encountering an existing well is $P(G \leq 0)$ or the integral of equation 2.6 from $-\infty$ to 0.

$$G = S - [F + D] \quad (2.6)$$

The convolution integral, equation 2.7, is used to sum two independent random variables, $Z = F + D$, and is applied to the bracketed terms in equation 2.6.

$$p_z(z) = \frac{d}{dz}p_z(z) = \int_{-\infty}^{z} p_f(f)p_d(z-f)df \quad (2.7)$$

Similarly, the cross-correlation integral is used to find the difference of two independent random variables, $G = S - Z$, given in equation 2.8.

$$p_g(g) = \frac{d}{dg}p_g(g) = \int_{-\infty}^{g} p_z(z)p_z(g+z)dz \quad (2.8)$$

Using arithmetic on probability distributions in this manner is similar to the distributions generated from load and resistance factor design (LRFD) in structural mechanics. The integral of the resistance minus the load in LRFD from $-\infty$ to 0 provides the probability of a load exceeding a material’s resistance. Similarly, integrating $p(g)$ from $-\infty$ to 0 in equation 2.8 provides the probability that a well will exceed the depth of the fractured region.

### 2.3.3 Total probability of encounter

The two previous subsections have provided the probability of a new horizontal hydraulically fractured well encountering an existing well in the horizontal plane only, and the vertical plane only. The total probability in three dimensions, $P_{int}$, is the multiplication of $P(G \leq 0)$ and $P_{h}$, shown in equation 2.9.

$$P_{int} = P(G \leq 0) \ast P_{h} \quad (2.9)$$
It is important to note that $P_{int}$ is not sufficient to establish the risk of communication to the surface or shallow aquifers because it does not incorporate the probability that an existing well provides a pathway for fluids to migrate upwards in, rather it is a necessary condition.

### 2.4 Sensitivity Analysis

The above analytical approach has been applied to the New York State well data set. Since there are four user supplied parameters that can influence the final result, we examine the importance of each via a sensitivity analysis. The four input parameters are:

1) The area of influence of each hydraulic fracturing job ($A_{hf}$) which we allow to range from 1,500 ft$^2$ to 1,600,000 ft$^2$.

2) The number of wells included for each calculation point ($n$) which ranges from 20 wells to 120 wells.

3) The inclusion of vertical fracture growth ($F$) to the model.

4) The depth at which the horizontal segment of the fracturing well is located, where fracture growth is initiated ($S$).

Values for $A_{hf}$ were chosen based on typical well sizes and assumed fracture growth and orientation. Horizontal wells within the Marcellus region typically have lateral lengths between 2,000 ft. and 4,000 ft. and a well pad can support multiple wells (between 6 and 8). $A_{hf}$ is therefore allowed to range from 1,500 ft$^2$ to 1,600,000 ft$^2$. The upper value of 1.6 million ft$^2$ is used as an upper bound and assumes that fractures are complex and create permeable zones around the well that extend 400 ft. in either horizontal direction perpendicular to 8 wells, for 1/8 of each of their 4,000 ft. length. The lower
bound of 1,500 ft\(^2\) assumes that fractures are not complex and instead have a width of 1/4 inch extending 400 ft. in either horizontal direction perpendicular to 6 horizontal wells. The wells each extend 3000 ft. and have a single fracture every 200 ft. along the length.

The measure selected to assess the sensitivity of each parameter was the average probability of encountering an existing well with the installation of a new HF well across the entire region considered (the Marcellus Shale within New York). The average of the resultant probability \(P_{int}\) across all representative areas provides a metric to easily compare each model configuration.

The model was run using ten values of \(A_{hf}\), eight values of \(n\), and three strategies for the HF well depth within the shale. To test the impact of fracture growth the model was also run with and without fractures. Taking all combinations of these parameters yields 320 different parameter configurations.

Figure 2.7 shows the average probability across \(A_{hf}\) holding \(n\) constant at 60 wells (top pane) and the average probability across \(n\) holding \(A_{hf}\) constant at 12,000 ft\(^2\) (bottom pane). This provides insight into the sensitivity of each parameter, which can be examined in more detail by looking back at the equations that form the probability estimate. Increasing \(A_{hf}\) increases \(P_{h1}\) in equation 2.3, which in turn increases \(P_h\) in equation 2.4. The impact of adjusting \(n\) is more difficult to assess through evaluating the equations, because \(n\) plays a role in the horizontal probability of encountering an existing well through equations 2.3 and 4. An increase of \(n\) causes \(P_h\) to increase, but decreases \(P_{h1}\) by making \(A_{RA}\) larger. The direct impact of these changes is difficult to assess analytically since the rate at which \(A_{RA}\) changes with \(n\) is different for each calculation point. Additionally, increasing \(n\) increases the number of wells that contribute to \(D\), which may cause
the vertical probability of encounter to shift slightly. It is found that increasing the number of wells used for each calculation point does little to the average probability when compared to the other parameters. The difference between using \( n=20 \) wells and \( n=120 \) wells is very small, at most increasing the average probability by 0.27\%. The sensitivity of \( n \) becomes more apparent when looking at discrete calculation points rather than the average probability of all calculation points.

To assess the importance of the fracture propagation three scenarios were run, 1) when vertical fracture propagation was not included and the depth to the horizontal fracturing well was set to the base depth of shale in the region, 2) when vertical fracture propagation was not included and the depth to the horizontal fracturing well was set to the top depth of the shale in the region and 3) the depth to the horizontal fracturing well was set to the base depth of shale in the region and fracture propagation was included (this represents a more realistic and conservative scenario in which wells are drilled so that they are near the bottom of the target formation). The use of \( F \) predictably increases the average probability values over not using \( F \), but interestingly setting \( S \) to the top depth of the Marcellus Shale without any fracture propagation produces the highest average probability values. This means that the fractures are rarely propagating to formations above the Marcellus Shale based on the statistical distribution of fracture heights from figure 2.1. Disregarding fracture propagation and setting \( S \) to either the top or bottom of the shale therefore provides a reasonable upper and lower bound on the average probability. Fractures do propagate and we will therefore accept that setting \( S \) to the bottom of the Marcellus Shale and allowing upward fracture propagation to occur provides a reasonable estimation of the probability of a hydraulic fracturing job encountering an existing well.
Disregarding fracture propagation will effectively eliminate some existing wells that are above the target formation, but still close enough to have a non-zero probability of intersecting hydraulic fractures. This means that the probability of individual RAs may be smaller when disregarding fracture propagation. Consider the scenario where an RA contains existing wells that are all above the target formation, but still close enough to be within the region where fractures may grow to. Disregarding fracture growth and setting the $S$ to the top of the target formation will produce a probability for that RA of 0. However, setting $S$ to the bottom of the target formation and allowing fracture propagation will yield a non-zero probability since those wells are deep enough to intersect some of the longer hydraulic fractures. This will vary based on the thickness of the target formation, making the use of fracture propagation important when looking at the probability results spatially, while the average probability provides an easy way to compare different parameters in a global sense.
Figure 2.7. Average probability of model configurations holding $n$ constant at 60 wells (top pane) and the average probability of model configurations holding $A_{hf}$ constant at 12,000 ft$^2$ (bottom pane). The three lines refer to where fractures were initiated from, the top or bottom of the target formation, and whether upward fracture growth was included.

The non-averaged data must be examined to take into consideration spatial variability in $P_{int}$ so that regions of higher and lower probability can be identified. Figure 2.8 is a representation of the probability of encountering an existing well at all 8873 RAs for $A_{hf}$ set to 12,000 ft$^2$ and $n$ set to 20, 60, and 120 wells. In all cases the HF well is placed at the base of the Marcellus Shale and fractures growth is included. The western region is an area containing the highest density of existing wells that reach a depth that is equal to or
greater than the base depth of the Marcellus Shale. This high density of deep wells produces a region in which the probability of encountering an existing well is comparably high across all parameters. This region, however, contains less thermally mature gas making it less likely to be targeted by hydraulic fracturing companies as evident by the pattern of drilling just across the state line in Pennsylvania. The eastern region of the Marcellus Shale within New York is the most likely area to be targeted for hydraulic fracturing based on the trend of well placements occurring across the border in Pennsylvania (PA DEP, 2015, unpublished data from www.depgis.state.pa.us/PaOilAndGasMapping/). At the point scale, the impact of increasing $n$ can be seen as smoothing the probability across neighboring RAs. This is because RAs will share a greater number of existing wells with their neighbors. The effect is most evident in the blurring of the western region of the Marcellus Shale in figure 2.8. To more closely analyze the impact of changing parameters, the standard deviation was calculated for $A_{hf}$ and $n$.

The standard deviation of the probability results over a single parameter allow for areas of higher variability to be identified. Figure 2.9 shows the standard deviation of the probabilities for all values of $n$ with $A_{hf}$ held constant at 12,000 ft$^2$ in the top plot and the standard deviation of the probabilities for all values of $A_{hf}$ with $n$ held constant at 60 wells in the bottom plot. From these figures the increased variability associated with changing the hydraulic fracturing area of influence is apparent, with some areas having a standard deviation of as much as 13%. However, the number of points with a standard deviation exceeding 5% is 573 calculation points or 6.5% of the calculation points.
The standard deviation associated with changing the number of wells used for each representative area has a much smaller maximum than varying $A_{hf}$, with a maximum standard deviation of 0.85%. Hotspots in standard deviation for varying $n$ are the results of the smoothing seen in figure 2.8. Too many wells in the same region creates a regional average that washes out high local probabilities seen when a smaller $n$ is used. This can be seen most evidently in the north-central region of the Marcellus Shale in figure 2.8, where the white area fades as the number of wells used per calculation point is increased. Higher standard deviations for varying $A_{hf}$ are primarily due to the regional well density of an area because of the role that $A_{hf}$ plays in equations 2.3 and 2.4.

2.5 Results

The model indicates that the few hundred feet of upward fracture growth has very little impact on the probabilities calculated when compared to the plan view area that a new HF well takes up. Furthermore, the probability of encountering an existing well is on average lower for wells that have fractures initiated from the base of the target formation than those cases tested in which wells were not vertically fractured and placed at the top of the target formation. This indicates that fractures growing out of the target formation does not increase the average probability more than setting $S$ to the top of the formation with no fracture growth. Comparing each RA individually shows that 2282 out of 8873 RAs (25.7%) have probabilities that are lower by at least 0.01% when $S$ is set to the top without fracture growth compared to the bottom with fracture growth. The use of fracture growth is important to the results of individual RAs, while not producing values that are much above those probabilities calculated using the top of the target formation. We find
that only 1 RA had a probability value that was more than 0.8% lower (the difference was 2.53%) when $S$ was set to the top of the target formation rather than the bottom with fracture growth.
Figure 2.8. Percent probability of encountering an existing well with a new horizontal hydraulic fracturing job in New York State underlain by the Marcellus Shale. HF area of influence is 12,000 ft$^2$ and the number of wells in each calculation point is 20, 60, and 120 respectively. Contours are delineated at the color bar labels.
2.6 Discussion

The number of actual well encounters is difficult to determine since not all encounters will result in communication; however, records on interwellbore communication are reported. Interwellbore communication has been documented in Alberta by the regulatory agency Energy Resources Conservation Board, now Alberta Energy Regulator. Reported interwellbore communication has occurred 20 times from 2009 to 2012 in Alberta where multi-stage fracturing wells connected to existing wells (Kim, 2012). Of
these 20 incidents, 55% resulted in no long-term adverse effects on production and commu-
nication was within the same formation. The maximum distance of communication
was 2400 m, the closest was 30 m, and the median was 250 m. During the same time pe-
period 5041 multi-stage fracturing wells were constructed. This leads to a known interwell-
bore communication rate of 0.4%. In addition, 95% of those incidents were horizontal to
horizontal well communication. That means that the probability of encountering a vertical
well was 0.02% of monitored wells in Alberta. All impacted wells were active at the time
of communication, where monitoring would likely be focused. Abandoned wells are not
instrumented to determine if interwellbore communication has occurred in those instanc-
es where there is no impact at the surface.

A 2010 report by BC Oil & Gas Commission states that as of 2010 18 fracture
communication incidents had occurred in British Columbia (BC Oil & Gas Commission,
2010). The report states that these incidents primarily involved horizontal to horizontal
communication and ranged in distances from 50 m to 715 m. These reports led to the rec-
ommendation that operators cooperate through notifications and monitoring of all drilling
and completion operations where fracturing takes place within 1000 m of existing well-
bores.

One specific incident in Alberta led to hydraulic fracturing and formation fluids
being released at the surface of a nearby horizontal producing oil well (Energy Resources
Conservation Board, 2012). The wells were 129 m apart at the closest point and were
both targeting the same formation. The communication was deemed to be a result of not
maintaining the company’s protocol of spacing wells a minimum of 135 meters apart.
The number of cases in Alberta and British Columbia shed light on the reality of inter-wellbore communication even with well documented existing wells and surveys conducted by operators.

This study has found that certain regions of New York underlain by the Marcellus Shale have probabilities on the order of 1-2% of encountering existing wells, while the vast majority of regions have much lower probabilities. The average probability for this entire area is found to be around 0.025% when $A_{hf}$ is 12,000 ft$^2$. An area of 12,000 ft$^2$ represents a cluster of ten $\frac{1}{4}$ inch fractures extending approximately 250 feet in plan view in both directions every 200 feet from a horizontal well that is approximately 4000 feet long for a pad containing 6 horizontal wells. Due diligence on the part of oil and gas well operators is the largest factor in reducing the probability of new hydraulic fracturing wells encountering an existing well. Surveys of existing wells, their depths, and closure status will reduce the probability that new wells encounter existing wells. This study therefore provides insights into areas where more care should be taken if new horizontal hydraulic fracturing wells were to be constructed in the Marcellus Shale in New York.

2.7 Conclusion

Encountering an existing well does not guarantee that a wellbore is capable of providing a pathway to the surface. Potential pathways include: cracks between cement and surrounding rock formations, spaces between the cement and well casing, voids created by well casing corrosion, and pathways through the interior of the well itself. Pathway development is dependent on whether a wellbore has failed, and failure is dependent on the quality of construction, lifetime monitoring, and the care taken during shutin
(Wojtanowicz, 2008; Bachu and Watson, 2009; Watson and Bachu, 2009). The probability of encountering a well is the first step in assessing the risk of a hydraulically fractured well communicating with shallow aquifers, and places an upper bound on that risk. Further work on assessing the probability of an encountered well providing a pathway needs to be done to determine the risk that fracturing poses to shallow aquifers.

2.8 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1247437. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors acknowledge the Vermont Advanced Computing Core which is supported by NASA (NNX 06AC88G), at the University of Vermont for providing High Performance Computing resources that have contributed to the research results reported within this paper.

Well data for New York State is available from the NY DEC at http://www.dec.ny.gov/energy/1603.html.

The authors thank Richard Davies and Simon Mathias from Durham University for graciously providing induced fracture propagation data from shale gas formations throughout the United States.
Additionally, the authors acknowledge the contributions of Patrick Fry for his work in consolidating and cleaning several of the data sets; and of Eric Hernandez for his comments throughout the revision process.

2.9 References


BC Oil & Gas Commission (2010), Communication during fracture stimulation, BC Oil & Gas Commission, Victoria, British Columbia.


Leff, E. (2011), Revised draft supplemental generic environmental impact statement on the oil, gas and solution mining regulatory program, New York DEC.


Nordbotten, J.M., M.A. Celia, S. Bachu, and H.K. Dahle (2005), Semianalytical solution for CO₂ leakages through an abandoned well, Environmental Science and Technology, 39, 602-611.


CHAPTER 3 Predicting Gas Migration through Existing Wells

James A. Montague¹
George F. Pinder¹
Theresa L. Watson²

1. Civil and Environmental Engineering, University of Vermont, Burlington, Vermont, USA.
2. T.L. Watson & Associates Inc., Calgary, Alberta, Canada

3.1 Abstract

The ability to accurately predict the probability of fluid migration from depth through existing wells based on known well properties, such as age and depth, would be enormously helpful in understanding how migration pathways develop and the identification of potential migration without extensive field tests. The presence of fluid pathways is an important environmental issue because such pathways allow gas, either naturally occurring methane or sequestered CO₂, to move into the atmosphere. In this paper, we explore the ability of various predictive models to forecast gas migration at existing wells in Alberta, Canada based upon the characteristics of existing deep wells. Alberta was selected as a case study because of the availability of data in an area that has required wells to be tested for pathway development after rig release since 1995. Wells that do not demonstrate pathway development require no further testing until the well is abandoned. We show that accurately predicting fluid migration requires detailed information on well construction, production, and fluid properties; and even then, the models considered in
this study misclassify a large number of wells. This suggests other factors may contribute to pathway formation. Of the models investigated, random forests provide the best results on this data set, correctly identifying 78% of the wells used.

3.2 Introduction

There are a large number of existing deep wells in many oil and gas producing regions. In Alberta Canada, there are over 300,000 wells (Watson and Bachu, 2009), and it is infeasible to test all abandoned wells as potential conduits to the surface in a cost-effective and efficient manner. With relevant well properties, we are able to predict which wells are likely to provide a pathway for gas migration; to this end, the large database of well characteristics provided by IHS AccuMap is useful in training the predictive models.

3.2.1 Gas Migration

Methane is generated naturally by organic material breaking down either through microbial methanogenesis or through temperature and pressure dependent thermogenic processes at depth (Osborn et al., 2011). Differentiating between these processes of formation provides insight into where the methane originated. Methane is classified as either thermogenic or methogenic using the ratio of stable carbon isotopes, known as $\delta^{13}$C, and the ratio of methane to higher-chain hydrocarbons. The ratio, also known as the carbon fingerprint of a sample, is used to approximate the source depth of methane samples.

The extent of gas leakage through wells was examined by Kang et al. (2014) in McKean County and Potter County, PA. They measured methane emissions from 19
wells as well as 19 control points located near the measured wells. Control points were chosen in various land covers to be consistent the land cover where wells were sampled. Methane flow rates that exceeded the flow rates at the control locations were observed in all 19 wells. Carbon fingerprinting determined most wells produce some thermogenic methane, suggesting deeper sources than those associated with the control samples.

The Alberta Energy Regulator (AER) keeps a record of methane emissions from energy wells within the province. AER identifies two general forms of emissions that can occur after a well has been installed: gas migration (GM) or surface casing vent flow (SCVF). Both GM and SCVF indicate a pathway is present from a source formation along the well casing. GM is defined as the flow of gas that is detectable at the surface external to the outermost casing string (Alberta Energy Regulator, 2003). On the other hand, SCVF is the flow of fluid in any casing annuli of the well but typically the annulus between the surface casing and production casing. The AER requires testing for GM and SCVF at all wells within a designated area in southeastern Alberta since 1995; we will refer to this region as the test area, see figure 3.1. Testing occurs at the time of rig release and when the well is abandoned. If a well does not demonstrate GM/SCVF at the time of abandonment, no further testing is required; however, if a well exhibits GM/SCVF further testing is required every year for 5 years. The seriousness of the leakage determines the actions that are required to be taken to remediate the problem.
Inspection for SCVF can be as simple as a bubble test, in which a small hose is connected to the surface casing vent and immersed in 1 inch of water and monitored for bubbles. GM is generally measured with gas detectors at the soil surface, but was traditionally done by auguring small holes in the soil and testing with a lower explosive limit meter. More advanced methods capture gas in sealed containers so that a flow rate in m$^3$/d or kg/d can be measured and the gas analyzed (Kang et al., 2014).

3.2.2 Factors that Contribute to the Formation of Pathways Within the Annulus of Deep Wells

Predicting the probability of GM and SCVF depends upon knowledge of the characteristics of relevant wells. The method of construction for energy wells is the primary factor contributing to pathways development. This includes a variety of factors such as
the type of cement used, the length of the annulus that is filled with cement, the length of
the well that is cased, and the deviation of the well from vertical. Other factors, such as
the geographical location of the well, play a role in whether GM/SCVF has occurred or
will occur (Bachu and Watson, 2009). Age may also contribute to the degradation of ce-
ment within the well, which can lead to pathways forming years after a well has been
plugged and abandoned (Dusseault, Gray, and Pawel, 2000).

While the aforementioned factors may contribute to the formation of pathways in
a given well, as Watson and Bachu (2009) point out, the formation of pathways is a large-
ly random process that is difficult to predict. In order to determine whether a pathway
will form or has formed, predictive models are employed that use well and fluid proper-
ties to produce a binomial result; either the well has provided a pathway or it has not pro-
vided a pathway.

3.3 Training Data

The data on well properties and locations that constitute the inputs to our predic-
tive models were provided by IHS AccuMap, a company which offers a software tool to
query a large database of energy wells within Alberta. Data that was extracted and used
included the physical properties of the wells, construction properties, operational data,
fluid data, and spatial data. The physical properties include the total depth of wells, the
depth of the completed interval, the deviation from vertical of the well, and the casing
diameter. The construction properties include the spud date (date in which drilling be-
gins) and the time to rig release. The operational data include the number of wells per
operator (a surrogate for the size of the operator), the production time in hours, the cumu-
relative fluid produced, and the number of years the well has been abandoned. The fluid data include the gross heating value (GHV), vapor pressure, molecular weight, relative density, critical pressure, and critical temperature. And finally, the spatial data include the locations of wells and the density of wells in the area (wells/mi$^2$).

The wells used to train the various models are all located within the test area in southeastern Alberta. This choice was made to ensure every well used in this analysis is regularly monitored for GM/SCVF. The test area contains 38,391 wells; of these wells, only 4,024 have complete property data. The remaining wells are missing at least one of the properties mentioned above. Figure 3.2 shows the number of wells that are missing each property. The majority of missing data describe fluid properties; these six variables represent the chemistry of the fluid extracted from the wells. Removing the fluid properties from the dataset increases the number of wells with completely observed properties to 28,534 (74.3%).
3.4 Methods

Seven predictive models were used to predict GM/SCVF using the same set of wells located in the test area of Alberta, Canada. Each model used an identical training set and validation set so the results could be directly comparable. Two challenges arise when training the models and assessing their performance. The first challenge stems from missing property values within the data set and the second challenge is from the unbalanced classification of the data between wells that have experienced GM/SCVF (class 1) and wells that have not been reported to experience GM/SVCF (class 2). Some class 2 wells may, in fact, experience GM/SCVF but since testing is not required after a well has had a negative test for GM/SCVF. Additionally, wells constructed prior to 1995 may not have been tested for GM/SCVF.
The predictive models used require complete properties for each well used for training and validation. Several strategies exist to handle missing properties within a data set. The easiest solution is to train and predict on only complete data sets. This ensures that all utilized properties have real observations, but it also reduces the number of observations used for training which may eliminate valuable information (Horton and Kleinman, 2007).

Many wells within the Alberta wells data set contain at least one missing property, the percentage of wells eliminated from the data set by the elimination of incomplete wells is approximately 90% of the measured wells. Eliminating 90% of the wells from training limits the variety of wells that can be used for training and thus limits the predictive power of models. It is thus desirable to use all wells in which information has been collected.

Wells with incomplete property values require some value to be assigned to replace the missing value. There are several possible methods to replace missing values, the simplest option is to replace all missing values with the same value, such as 0. This creates a large number of wells that are artificially closer together for a given property. Slightly more sophisticated methods involve using the mean of all available information to impute missing values, or imputing values based on the mean of the k-nearest neighbors. A more time-consuming approach is to develop a model to predict missing properties based on those that are available (Horton and Kleinman, 2007). In order to utilize all available wells, we chose to impute data using an iterative principal component analysis (PCA) fitting method called trimmed scores regression (TSR) (Artega and Ferrer, 2002).
Trimmed scores regression follows an iterative process in which an initial PCA model is fit to the mean centered data. Missing properties are initially replaced by 0, but in each subsequent iteration, the missing properties are imputed using the interpolation from the PCA model. This process is repeated until the imputed values converge to below a given threshold (Folch-Fortuny, Artega, and Ferrer, 2015; 2016). Using this method of imputation takes advantage of the known properties for each well to predict the missing properties.

Addressing the second challenge, that is the unbalanced classification of data between wells, requires that assessment metrics can handle unbalanced classifications. The available data for training and validation contain primarily class 2 wells that have not experienced GM/SCVF, making up 87.35% of the total wells. Highly skewed classification in training tends to push models towards classifying the large number of class 2 samples with high accuracy, but classifying the small number of class 1 samples with low accuracy (Kubat et al., 1998). Solving the problem of skewed data either requires the samples selected for training be artificially balanced or the predictive model uses a metric for assessing performance that is independent of the number of samples in each class.

3.4.1 Model Performance

Assessing the ability of the above models to properly classify wells is done using a confusion matrix which represents reality as compared to model predictions; an example is shown in figure 3.3. Confusion matrices present the results from a single model run in an efficient manner to easily assess performance. However, to differentiate model
performance over many runs, it is desirable to have a single value representation of the quality of each classification.

The most commonly used metric of assessment is accuracy which is defined as the number of samples classified correctly over the total number of samples. Accuracy works well in assessing performance when the number of observed positive and negative samples are approximately equal. However, accuracy becomes less informative when one category contains many more samples than the other (Kubat et al., 1998; Valverde-Albacete and Peláez-Moreno, 2014).

To illustrate this concept, consider a classification task in which there are 90 negative samples and 10 positive samples. Classifying all samples as negative yields an accuracy of 0.9 (i.e. 90/(90+10)). An accuracy of 0.9 may seem like a good result, but the model is doing a bad job if you want to predict positive results- it has a sensitivity of 0 (i.e. 0/10).

Other assessment metrics attempt to overcome this by using the probability of randomly selecting the correct class based on the population, or by weighting based on the true populations of each class. We chose to use the geometric mean (g-mean) between sensitivity and specificity to better understand the performance of the models used (Kubat et al., 1998).
3.4.2 Predictive Modeling

Predictive models are used to calculate a response based on one or more inputs. These models take many forms and can range from simple linear combinations of inputs to much more complex mathematical relationships. We evaluate logistic regression models, support vector machines, random forests, and back-propagation neural networks for their ability to predict the occurrence or future occurrence of GM/SCVF based on the characteristics of each well. A brief description of each of these models follows.

Logistic regression (GLR) is the simplest and most transparent method we consider; it takes a linear combination of input parameters and predicts a binomial response (i.e. does the well experience GM/SCVF or not?). The linear combination of parameters is passed through a logistic function which yields the probability of a positive response (Casella and Berger, 2002). This method is commonly used in medical diagnosis (Boyd et al., 1987) as well as determining the probability of failures in engineered parts and systems (Palei and Das, 2009).
The term artificial neural networks encompass a large range of predictive models that are inspired by our understanding of how the brain learns through rearranging and strengthening connections between neurons (Negnevitsky, 2005). Several distinctly different neural networks are used in later sections of this paper: these are back-propagation networks, a random forest, a support vector machine, and a generalized regression network.

Back-propagation networks use a simple computational architecture in which input vectors are connected to a randomly initialized hidden layer which, in turn, is connected to an output layer. Each layer is fully connected by a matrix of weights. Weights are iteratively adjusted by propagating the error between the network output and the true output back through each layer using gradient descent (Rumelhart et al., 1986). Iterating until the error stabilizes or drops below a threshold produces a trained network that can be used to predict responses to new inputs. The basic back-propagation network has been refined to get out of local minima and to converge faster using momentum and adaptive adjustment of the learning rate (Jacobs, 1988; Negnevitsky, 2005). The models used in this paper include back-propagation with Bayesian regularization (BR), back-propagation with variable learning rate gradient descent (GDX), and back-propagation with Levenberg-Marquardt optimization (LM).

Random forests are a collection of decision trees each trained on a random subset of features. Each tree is trained to classify the data with near perfect accuracy and with a large enough collection of decision trees random forests are capable of generalizing well (Ho, 1998). This means, unlike other classification methods, random forests improve their ability to generalize as their complexity increases. Individual decision trees are
trained on a small subset of features. Beiman (2001) describes a method for growing forests that allows for the importance of each feature to be measured, thus giving insight into the mechanisms that have the greatest contribution to the classification.

A support vector machine (SVM) is a form of neural network that utilizes training data to form a discriminating hyperplane that separates data with the maximum possible margin between data classifications. By utilizing a small number of support vectors relative to the total number of available training vectors, support vector machines can achieve a high level of generality (Cortes and Vapnik, 1995). The support vectors work similarly to the first layer in a back-propagation neural network which passes weights to a classifier that determines which side of the hyperplane a new vector is on. Non-separable data requires the use of a soft margin hyperplane which minimizes the number of misclassified data points while also maintaining a high level of generality using a cost function (Press et al., 2007).

The final network used in this paper is the general regression neural network (GRNN) developed by Donald Specht (1991). GRNNs use a small subset of the data to approximate a response variable. A single pass of the training data fully trains the network, but iterative fitting is required to find the optimal sample probability width, $\sigma$. Larger values of $\sigma$ smooth the curve, while $\sigma \rightarrow 0$ forces the GRNN into a step function. To fit $\sigma$, a mean squared error between the predicted responses and the actual responses is calculated by successively holding a single training vector out of the training step. This ensures an artificial minimum error caused by a step function is avoided.
3.5 Improving Classification

The seven models described above along with randomly guessing were used to test how well the trained models could predict the number of wells exhibiting GM/SCVF. The results for the models are improved by adjusting the training data and changing how missing values are imputed. The models were initially trained on 80% of the wells, while 20% were set aside for validation. Training and validation sets maintain the same distribution of class 1 to class 2 wells as the overall data set.

The results of the first training are shown in figure 3.4. The various models are indicated on the horizontal axis and the assessment measure for each model is presented on the vertical axis. Each measure ranges from 0 to 1 where values closer to 1 indicate better model performance and values closer to 0 indicate worse model performance. Each measure is represented by a unique column for the model shown on the x-axis. Models that require random initialization of weights include error bars over 30 runs, while models that do not require random initialization have only a single point.

The results shown in figure 3.4 indicate most models can classify almost all class 2 wells, indicated by specificities of approximately 1. The high specificity is an effect of training on a dataset with highly unbalanced classifications (more class 2 than class 1) for most of the models used. The unbalanced training set forces models to value correctly classifying the large number of class 2 wells over the small number of class 1 wells. Random forests are immune to this since they generally fit training data perfectly when allowed to be sufficiently deep (Breiman, 2001). Training can be improved by adjusting the training sets to be more balanced.
Balanced training data causes methods that use gradient descent to penalize misclassification of both classes equally or approximately equally. The training set was changed to balance the examples from each class, or nearly balance the examples. The SVM was selected to illustrate the impact of varying the class ratios. The effect of varying the ratio of the two classes (class 1: class 2) between 1:1/5 and 1:5 is considered. In each case, the number of class 1 wells with GM/SCVF was kept constant (the
Figure 3.5 shows the result of those runs in obtaining a correct GM/SCVF identification for the validation data. The horizontal axis is the numerator in the ratio between the two classes. The point where the specificity and sensitivity lines cross shows where the model performs equally for the two classes. The value of 1:1.25 was therefore selected as the ratio of the two classes. This means there are still slightly more wells that have not experienced GM/SCVF in the training data.

Figure 3.5. Comparison of SVM performance over multiple ratios between the two classes. 1 represents an equal ratio of classes, <1 indicates more wells without GM/SCVF in the training data, and >1 indicates more wells with GM/SCVF in the training data.

Adjusting the ratio of wells within the training set of each classification has the effect of increasing the importance of one class over the other in how the individual models assess performance. A balanced ratio, or nearly balanced ratio, between the classes causes the optimization method of each model to value each class equally, or nearly equally. This can be seen in figure 3.6 when comparing the training results of the models trained using the intrinsic ratio of the data and those trained using 1:1.25 ratio. Figure 3.6
is made by subtracting the assessment metrics of the models trained on a ratio of 1:1.25 between class 1 and class 2 from the intrinsic ratio between class 1 and class 2. Positive values indicate the given assessment metric is higher for the intrinsic ratio between the classes, while negative values indicate the assessment metric is higher for a ratio of 1:1.25 between the classes.

Figure 3.6. Classification assessment for each predictive model based on training data only. The training data is more evenly distributed between the two classes with a ratio of 1:1.25 for wells with GM/SCVF to wells without GM/SCVF.

The difference in sensitivities for each of the models indicates using a ratio of 1:1.25 between class 1 and class 2 causes the models to better identify the class 1 wells.
that exhibit GM/SCVF. The difference in g-mean indicates each model performs better when using an approximately even ratio between the two classes.

The next line of investigation to improve the model focuses on the large number of missing properties associated with the fluid extracted from wells. Recall that figure 3.2 shows the number of wells missing each property. The percent of wells missing properties associated with the extracted fluid is around 78%. The amount of data that is missing for these properties may be too much for our method of imputation to accommodate. To verify the fluid properties were adding useful information to the models, we reran the training and validation excluding the fluid properties. This reduces the number of properties from 17 to 11. The comparison of the validation assessment metrics is shown in figure 3.7.
Figure 3.7. Comparison of assessment metrics for validation data for models trained with and without fluid properties positive values represent metrics that are higher for models trained with fluid properties as inputs and negative values represent metrics that are higher for models trained without fluid properties.

Eliminating the fluid properties from the model inputs reduces the number of wells with incomplete data and reduces the number of imputations required for training; however, eliminating the information provided by the fluid properties causes a small drop in performance of some of the models. The random forest performs better with regards to all assessment metrics when using fluid properties. All other models can better identify class 1 wells when fluid properties are removed from the training set, shown in figure 3.7 by the sensitivity. Reducing the number of inputs for the models makes them less com-
plex and means additional data acquisition is not needed to provide full inputs to the models. Because so many of the wells are missing fluid properties, it becomes necessary to compare models trained with fluid properties and without fluid properties using only those wells in which all properties are observed. Comparison of a selection of trained models on the small subset of 4024 wells with complete properties is shown in figure 3.8. Models trained with fluid properties perform better than those trained without the fluid properties. This indicates that when present fluid properties provide valuable information for predicting GM/SVCF. The effect may be related to the stimulation, construction, and producing methods used depending on the targeted fluid. For instance, an oil well will have artificial lift whereas a gas well will not. Therefore, knowing the type of target fluid can help predict the stresses a well will experience over its producing life.
3.6 Discussion

The predictive models utilized for this analysis could perform better than random guessing when classifying wells in Alberta for GM/SCVF; however, they have their limitations. The best model achieves a g-mean of 0.68. The fluid flowing through wellbores plays a role in whether a well will experience GM/SCVF, whenever it is available fluid properties contribute to better well classifications. This seems to indicate that the fluid plays a role in the formation of pathways that allow for gas to migrate upwards.
The random forest model is the best performer of the models tested on this data set. It was able to classify 62% of the wells with GM/SCVF correctly and 74% of the wells without GM/SCVF correctly within the validation data set. The percentage of wells classified correctly using the random forest increases to 87% of the wells with GM/SCVF and 74% of the wells without GM/SCVF when only wells in which all properties are observed are used for prediction. The relative importance of each property can be determined during the construction of the random forest (Breiman, 2001). Figure 3.9 shows that the most important features in predicting GM/SCVF are the deviation of the well from vertical and the year in which the well was constructed, consistent with what Watson and Bachu (2009) found in their work. The least important property for prediction is the density of wells within a given area.

Figure 3.9. The relative importance of each feature as described by 30 runs of a random forest model.
Prediction of GM/SCVF is inhibited by imperfect data. The regulations in place within Alberta for testing GM/SCVF mean that wells that do not exhibit leakage at the time of abandonment require no further testing. This causes the years abandoned property to have very little impact on the predictive capabilities since wells are rarely tested years after they have been abandoned. Requiring continuous monitoring of wells within a small area, such as the test area identified in figure 3.1 would allow the models to be retrained with more accurate information and may provide insights into how the length of time abandoned affects leakage.

3.7 References


CHAPTER 4 Visualization and Simulation of Density Driven Convection in Porous Media using Magnetic Resonance Imaging

James A. Montague\textsuperscript{1}, George F. Pinder\textsuperscript{1}, Jay V. Gonyea\textsuperscript{2}, and Scott Hipko\textsuperscript{2}, Richard Watts\textsuperscript{2}

\textsuperscript{1}. Civil and Environmental Engineering, University of Vermont, Burlington, Vermont, USA.
\textsuperscript{2}. Department of Radiology, University of Vermont College of Medicine, Burlington, Vermont, USA.

4.1 Abstract

Magnetic resonance imaging is used to observe solute transport in a 40 cm long, 26 cm diameter sand column that contained a central core of low permeability silica surrounded by higher permeability well-sorted sand. Low concentrations (2.9 g/L) of Magnevist, a gadolinium based contrast agent, produce density driven convection within the column when it starts in an unstable state. The unstable state, for this experiment, exists when higher density contrast agent is present above the lower density water. We implement a numerical model in OpenFOAM to reproduce the observed fluid flow and transport from a density difference of 0.3%. The experimental results demonstrate the usefulness of magnetic resonance imaging in observing three-dimensional gravity-driven convective-dispersive transport behaviors in medium scale experiments.
Keywords: Magnetic Resonance Imaging; Porous Media; Numerical Modeling; Density Driven Convection

4.2 Introduction

Small differences in fluid density create complex fluid flow paths. These complex flow paths enhance fluid mixing and therefore dispersive transport. We visualize the concentration resulting from these complex flow paths using magnetic resonance imaging (MRI).

4.2.1 Density Driven Convection in Carbon Capture and Storage

An important problem wherein density driven convection plays a role is carbon capture and storage (CCS). This is because convection decreases the dissolution time of CO$_2$ in deep saline aquifers. Supercritical CO$_2$ injected into saline aquifers will rise to the top from the positive buoyancy of CO$_2$. Over time, CO$_2$ slowly dissolves into the top layers of brine, increasing its density (IPCC, 2005). The density increase in the top layer relative to the brine below gives rise to unstable gravity driven flow. Fingering develops as local convection moves denser CO$_2$ saturated brine downward and replaces it with the less dense brine. This phenomenon has been observed at laboratory scales in two-dimensional Hele-Shaw cells (Faisal et al., 2013; Kneafsey and Pruess, 2009).

Kneafsey and Pruess (2009) used a Hele-Shaw cell and the tracer bromocresol green to show experimentally that density-driven convection was occurring with fluid density differences ranging between 0.1% and 1%. Faisal et al. (2013) expanded upon
this work by varying the “permeability” of the Hele-Shaw cell by changing the width of the gap between the two glass plates. The variation in “permeability” showed that small changes greatly affected the time for fingering to reach the bottom of the cell.

4.2.2 NMR Measurements in Porous Media

Nuclear magnetic resonance (NMR) has been used for non-invasive imaging of the structure of porous media. Issa and Mansfield (1994) used $T_1$ weighted images to estimate permeability in sandstones using the relationship between the $T_1$ of porous media and the $T_1$ of water. Komlosh et al. (2011) determined the average porosity of glass capillary tubes on the voxel scale (three-dimensional volume over which measurements are averaged) using double pulsed-field gradient diffusion weighted MRI. Saturation curves for various porous materials as was calculated by Muir et al. (2014) using $T_2$ mapping Spin Echo Single Point Imaging as water was displaced by heavy water.

NMR has also been used to measure flow through porous media in small-scale experiments. Shattuck et al. (1995) measured thermal convection patterns using a spin echo technique in a small cylinder packed with monodisperse plastic beads. Kimmich et al. (2001) used NMR to map fluid velocities in percolation clusters created with a circuit board plotter. Khrapitchev et al. (2002) tracked the locations of particles in a small porous media column using velocity exchange spectroscopy to examine the dispersive flow within a column. This NMR method allowed them to track molecules by applying a spin and observing their displacement after a small delay in a 10 mm diameter column. Seymour and Callaghan (1997) provide a good overview of the NMR methods used to image fluid flow in porous media.
In addition to mapping physical properties of porous media, the concentrations of magnetic tracers such as Gd-diethylenetriaminepentaacetate (Gd-DTPA$^{2-}$) can be tracked through time. Haber-Pohlmeier et al. (2010) demonstrated that Gd-DTPA$^{2-}$ can be tracked within a sand column in an MRI scanner to determine flow and transport behaviors in three-dimensions.

Contrast agents are required to be compounds that change the relaxivity of the solution in which they are dissolved; alternative contrast agents include NiCl$_2$ (Pearl et al., 1993) and CuSO$_4$ (Greiner et al., 1997). Greiner et al. (1997) packed two small columns with glass beads and calculated the concentration of copper sulfate at the outlet with both MRI and atomic absorption spectroscopy. The results for both homogeneous and heterogeneous bead packing showed that MRI was able to accurately measure outflow breakthrough curves with the added benefit of quantifying the concentration within the column.

4.3 Materials and Methods

This work extends that of earlier research by focusing on the impact of low dye concentrations on gravity-driven convective-dispersive transport. More specifically, we use an intermediate scale physical model in conjunction with MRI and numerical modeling to visualize and evaluate the impact of gravity on porous-medium transport.

4.3.1 Experimental set up

The sand column employed in this work is designed to evaluate the impact of sharp interfaces on solute transport in porous media. The dimensions of the column are
designed specifically to make use of the maximum field of view of the MRI Center for Biomedical Imaging’s Philips 3T Achieva TX scanner located in the University of Vermont Medical Center. The scanner has a bore size of 60 cm and a usable volume diameter of 50 cm. Whole-body scanners are desirable for this study because they allow macroscopic flow to be captured over a large volume (Fukushima, 1999). Larger experiments also facilitate observation of scale-dependent behavior and limits the impact of column-wall boundary interactions on the experiment. A smaller magnet in either a horizontal or vertical configuration would allow for higher resolution concentration maps, but would suffer from limiting the size of soil strata used.

The column is packed with a centrally located cylinder of ground silica with a mean grain diameter of 10 microns. This cylinder is surrounded by clean Ottawa sand with a mean grain diameter of 200 microns. To realize this arrangement, the ground silica is first placed in a cylindrical sleeve, saturated, and frozen. A Shelby tube is used as the sleeve to freeze the silica for this experiment. This allows for easy extraction using a Shelby tube extractor. The fine silica cylinder is then cut to length and placed into the center of the large column. The annulus between the frozen silica is then filled with the clean Ottawa sand. After thawing, a sharp interface exists between the silica and the sand.

The basic schematic of the column is shown in figure 4.1. The column is built entirely out of plastic to allow safe imaging by the scanner. The presence of metallic objects or particles disrupts the magnetic field, resulting in signal loss, because MRI relies on a highly homogeneous magnetic field, both macroscopically and microscopically. Consequently, considerable effort has been taken to reduce the number of magnetic particles within our porous media by screening the sand with a strong neodymium magnet.
Figure 4.1. As built schematic of the plastic sand column.
4.3.2 MRI Scans

MRI uses the nuclear magnetic resonance signal from protons (hydrogen nuclei), in this case in water. The characteristic magnetic relaxation time of the longitudinal magnetization, $T_1$, depends on the intrinsic relaxation time of water in the medium, $T_{10}$, and the concentration of magnetic materials including the gadolinium-based contrast agent used in this study. The relaxation time is related to the contrast concentration by equation 4.1,

$$\frac{c}{M} = \left( \frac{1}{T_1} - \frac{1}{T_{1\text{back}}} \right) / r_i \quad (4.1)$$

where $c$ is the concentration of contrast agent in g/L in the saturated porous media, $M$ is the molecular weight of the contrast in g/mmol. The variable $T_{1\text{back}}$ is the background $T_1$ value of the porous media with distilled water in s$^{-1}$, and must be subtracted from the measured $T_1$ to remove the effects of different media on the measured signal. The variable $r_i$ is the relaxivity of the contrast agent; for the case of Magnevist, $r_i$ is 4.1 L/mmol.s based on empirical results (Tofts, 2010).

The spatial distribution of $T_1$ values can be calculated using various techniques. The spin-echo inversion recovery technique was used for this study. The recovery of longitudinal magnetization following inversion using a 180-degree RF pulse is described as an exponential function of the time between the inversion and the data acquisition (the inversion time, $TI$). It is given by equation 4.2,

$$M_z = M_0 \left( 1 - 2e^{-TI/r_i} \right) \quad (4.2)$$
where $M_z$ is the longitudinal magnetization, $M_0$ is the maximum magnetization, $TI$ is the inversion time, and $T_1$ is the unique value that describes the magnetic properties of each voxel.

Following the inversion time, $TI$, imaging data is acquired using a spin echo acquisition, resulting in an image in which the signal at each voxel is directly proportional to $M_z(TI)$. Note that this equation is based on the assumption of full recovery of magnetization between measurements, such that the repetition time $TR >> TI$. This assumption may not be justified for zero or very low concentrations of contrast agent, thus introducing a small bias in these cases.

Images are acquired with different values of inversion time, $TI$. Inversion times were 50, 100, 200, 300, 400, 600, and 800 ms. All other parameters are held constant, including the acquisition matrix (200x200), field of view (400x400 mm$^2$), slice thickness (10 mm), echo time (6.3 ms), and repetition time (1000 ms). The acquisition duration for each inversion time takes 3 minutes 23 seconds. At each volume element (voxel), the signal as a function of $TI$ is fit to the exponential function given in equation 4.2 to develop $T_1$ maps at each time point.

It should be noted that the range of concentrations that the MRI is sensitive to is small and consequently the density difference that can be studied is small. Concentrations of a gadolinium based contrast agent in the form of Magnevist ranged between 0 and 2.9 g/L. Concentrations of Magnevist exceeding 7 g/L cause fitting equation 4.2 to become difficult since at this concentration the $T_1$ value is smaller than the minimum inversion time of 35ms that we can measure with this technique.
4.3.3 Experimental Procedure

We initially fill the column with distilled water and scan it to get a baseline $T_1$ value for the sand and silica. Gadolinium contrast agent in the form of Magnevist is then pumped into the column for 2 hours and 10 minutes at a concentration of 2.9 g/L and flow rate of 100 mL/min until the coarse outer sand is fully saturated with contrast agent. This process is monitored using quick $T_1$-weighted scans in the sagittal plane (y-axis in figure 4.1, along which gravity acts). The fine-grained material remains free of contrast agent throughout pumping. After 2 hours and 10 minutes, pumping is stopped and the valves at either end of the column are closed. The column is then imaged using quantitative $T_1$ maps. While not being imaged, the column is stored horizontally in the same position. This is done to ensure that gravity is always acting along a single axis. The column is periodically imaged to observe the movement of contrast agent into the fine-grained material. After 40 days, the concentration of contrast agent equalizes throughout the column.

The column is then flushed with distilled water until the contrast agent is only present in the fine silica. Figure 4.2 shows the column before and after flushing with distilled water. The column continues to be imaged periodically for two months to show the effects of both diffusion and gravity on the concentration of Magnevist within the column.
4.3.4 Numerical Model

The concentration data collected from the MRI are used to develop a three-dimensional flow and transport model at the same resolution as the images produced by the scanner. OpenFOAM is used to model the observed physics by implementing a modified version of pisoFoam that includes mass transport and density driven convection. The model solves the Navier-Stokes equations with porous resistance in an iterative PISO (pressure implicit with splitting of operator) loop coupled to a convection-dispersion equation for mass transport that includes a time dependent Langmuir isotherm to partition mass between the fluid and solid phases. The transport equation is given in equation 4.3.

\[
\frac{\partial c^f}{\partial t} = \nabla \cdot \nabla c^f - \nabla \cdot uc^f - \frac{\partial c^s}{\partial t} \quad (4.3)
\]

In equation 4.3, the concentration is partitioned between the fluid phase, \(c^f\), and the solid phase, \(c^s\). The first term on the right-hand side of equation 4.3 is the dispersion term.
where $D$ is the dispersion tensor. It includes the effect of molecular diffusion and longitudinal and transverse dispersion components (Pinder, 2002). The second term on the right-hand side is the convection term where $u$ is the velocity vector. The change in the solid phase concentration is described by equation 4.4.

$$\frac{dc_s^x}{dt} = \left(1 - \frac{c^x}{c_{\text{max}}^s}\right)c_{\text{max}}^s a_{a,d}$$

(4.4)

Here, $c_{\text{max}}^s$ is the maximum concentration that can be in the solid phase as given by a Langmuir partitioning coefficient provided in equation 4.5. The adsorption speed of concentration moving from the fluid phase to the solid phase is $a_a$ and the desorption speed of concentration moving from the solid phase to the fluid phase is $a_d$. The adsorption/desorption speeds do not have to be the same and are in units of 1/T. Having differing coefficients ($a_a$ and $a_d$) represents cases in which a particle adsorbs quickly to a particle and forms a strong bond.

$$c_{\text{max}}^s = \frac{K_{La}bc_f}{1 + K_{La}c_f}$$

(4.5)

Equation 4.5 is the Langmuir isotherm, $K_{La}$ is the Langmuir coefficient and $b$ is the maximum adsorption capacity. Equations 4.4 and 4.5 together create a time dependent transfer of mass between the solid and liquid phases that asymptotically reaches the Langmuir partitioning between the two phases. The speed at which the asymptote is approached depends on the magnitude of $a_{a/d}$.

OpenFOAM was chosen for this work because it is an open source software that allows easy manipulation and addition to the included solvers (Weller et al., 1998). The code is robust and able to handle basic geometries, such as our column, with ease. Addi-
tionally, OpenFOAM’s PISO solver conserves mass and explicitly conserves momentum by correcting the velocity equation.

4.4 Results and Discussion

4.4.1 MRI Images

MRI scans were taken regularly in the sagittal and transverse planes. Each sequence is processed first into a $T_1$ map and then into a concentration image. Concentration images were made using equation 4.2 and then filtered to reduce noise within the image. The contrast agent was observed being transported out of the fine-grained material. Gravity driven convection was evident and was observed to reduce the time required for the contrast to move out of the fine-grained material. Figure 4.3 shows the progression of the contrast agent over time.
Figure 4.3. Sagittal plane concentration scan of sand column. Gravity is from top to bottom. The concentration of Magnevist is in g/L on a log-scale. A) 0 days after flushing B) 16 days after flushing C) 26 days after flushing D) 50 days after flushing. The color bar is presented in log scale to show better depict the plume of Magnevist coming from the fine-silica cylinder.

Figure 4.3A shows the contrast after initial flushing of the contrast agent resident in the coarse-grained sand with distilled water from the left side of the image. The low concentration tail to the right of the fine-silica cylinder in 3A occurs when flow paths around the fine-silica cylinder converge during pumping to create an area of stagnation.
The following panels in figure 4.3 show the contrast agent moving out of the fine silica under the effect of gravitational convection and dispersion. The gravity vector is from top to bottom in all panels of figure 4.3. Figure 4.3B shows the downward movement of the contrast agent after 16 days; it appears as the light-blue to white area below the fine-silica cylinder. In figure 4.3C, the leading edge of the contrast agent plume has reached the bottom of the column and begun to spread horizontally along the column. The final panel, figure 4.3D, shows that the concentration in the fine-silica cylinder has been reduced to 1 g/L from 3 g/L and that the concentration of contrast agent moving out of the fine-silica has also been reduced due to the smaller concentration gradient between the course sand and fine-silica.

The flow paths that the concentration takes in each panel of figure 4.3, as simulated by the calibrated model, is shown in figure 4.4. The closed system causes circulation to occur initiated by the unstable initial conditions of the fluid densities. The evolving pattern can be seen as the concentration moves out of the fine silica core and the difference in density is reduced. From figure 4.4A to figure 4.4D the peak velocity decreases by an order of magnitude indicating that the density is beginning to equilibrate throughout the column.
Figure 4.4. Sagittal plane model simulation of sand column. Gravity is from top to bottom. The concentration of Magnevist is in g/L on a log-scale. A) 0 days after flushing B) 16 days after flushing C) 26 days after flushing D) 50 days after flushing. The arrows represent flow direction at from the tail of the arrow. The color of each arrow represents the velocity in m/s on a log scale.

The measured $T_1$ value depends on the concentration of Magnevist as well as the surrounding material. We found that the scaling factor of 4.1 found by Tofts (2010) is accurate for converting $T_1$ values to concentration when Magnevist is in either pure water or the course sand. However, when the Magnevist is in the fine-silica porous media the
concentration calculated from the $T_1$ values requires a different scaling factor. We find that a scaling factor of 5.3 is appropriate for Magnevist in the fine-silica when comparing samples of known concentration.

The concentration calculated from $T_1$ values provides the total concentration; that is, both contrast agent that is in solution and contrast agent that has adsorbed onto the particles. MRI signals are formed by volume averaging over voxels meaning that each voxel will be influenced by the contrast agent that is in the solid phase and fluid phase.

### 4.4.2 Comparison of Simulations and Experiment

The numerical model was run with a three-dimensional mesh to simulate the impact of gravity. Figure 4.5 shows the line scan comparison for the images shown in figure 4.3 and the numerical model at the same point in time. Gravity is again acting from left to right in each panel of figure 4.5.

The model is able to capture both the peak in the fine silica and the lower concentration on the right side of each panel of figure 4.5. The peak concentration in the fine silica slopes in the direction of gravity because convection helps to move the concentration out of the fine silica. The model and experimental results have a root-mean-square error of 0.19 when comparing the experimental results to the modeled results at 16, 26, and 50 days.
Figure 4.5. Comparison line scans across the middle of the column for MRI results and model results. The concentrations for the model are broken into solid phase and liquid phase components. The sum of these two phases (black line) is comparable to the MRI line scan (blue circles). The concentration of Magnevist is in g/L. A) 0 days after flushing B) 16 days after flushing C) 26 days after flushing D) 50 days after flushing.

Some model parameters are known ahead of time and other parameters are determined by optimization. Table 4.1 shows the parameters used for modeling. The highlighted values are parameters found during initial optimization of the numerical model. Optimization is done using pySOT, an optimization algorithm designed to quickly and
efficiently search through parameter space when each computation is an expensive black-box function (Eriksson et al., 2015). The optimization is done over two regions, the course sand and the fine silica.

Table 4.1. Model parameters. Values correspond to the Course Sand and Fine Sand. Highlighted values indicate that parameters are fit using optimization.

<table>
<thead>
<tr>
<th>Sand Layer</th>
<th>Permeability (m$^2$)</th>
<th>Transverse Dispersivity ($\alpha_T$) (m)</th>
<th>Longitudinal Dispersivity ($\alpha_L$) (m)</th>
<th>Molecular Diffusion ($D_m$ (m$^2$/s))</th>
<th>Langmuir Coefficient ($K_{La}$ (m$^3$/kg))</th>
<th>Max. Adsorption ($b$ (kg/m$^3$))</th>
<th>Adsorption Speed ($a_a$ (s$^{-1}$))</th>
<th>Desorption Speed ($a_d$ (s$^{-1}$))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>6.78e-11</td>
<td>2.5e-5</td>
<td>2.5e-4</td>
<td>2.14e-11</td>
<td>1.98</td>
<td>8.22</td>
<td>6.06e-12</td>
<td>1.28e-11</td>
</tr>
<tr>
<td>Fine</td>
<td>1.56e-12</td>
<td>2e-6</td>
<td>2e-5</td>
<td>1.59e-11</td>
<td>7.16</td>
<td>0.86</td>
<td>3.31e-6</td>
<td>5.39e-10</td>
</tr>
</tbody>
</table>

The longitudinal dispersivity was set to the mean grain size and the transverse dispersivity was set to 1/10 of the longitudinal dispersivity (Haber-Pohlmeier et al., 2010). Molecular diffusion for Magnevist in water is reported to be approximately 4e-11 m$^2$/s by Osuga and Han (2004). The tortuosity of the porous media influences the diffusion of the Magnevist by increasing the number of obstructions with which molecules collide and bounce off. This has the effect of reducing the effective diffusion below the diffusion coefficient calculated in water.

Magnevist is reported to be a conservative tracer, meaning that adsorption should not occur (Haber-Pohlmeier et al., 2010). We, however, found that the model was unable to fit the observed results when adsorption coefficients were forced to 0. Furthermore, ICP-MS analysis of a batch adsorption test using the fine silica indicated that adsorption was not occurring when compared against the control. We suspect that there is an interaction within the fine silica that is retarding the movement of Magnevist in a way that mimics adsorption and desorption.
4.4.3 Model Sensitivity

The sensitivity of the model is evaluated by perturbing each parameter with a normal distribution defined by a coefficient of variation of 0.1, where the coefficient of variation is the standard deviation divided by the mean. The resulting coefficient of variation of the model output compared to the input variation gives a measure of how sensitive each parameter is to small perturbations. Figure 4.6 shows the average model coefficient of variation for each parameter compared to the input variation. For all variables, the resulting variation is less than the input variation indicating that the model is not sensitive to small perturbations to the inputs. Figure 4.7 shows the coefficient of variation for each voxel along the column’s cross-section from the center of the fine-grained silica to the edge of the course sand in the direction of gravity, the location of the line is shown in the inset of figure 4.7. Each point along the black line represents the coefficient of variation between model runs where the variable along the x-axis is perturbed using a normal distribution defined by a coefficient of variation of 0.1. The 50 perturbed samples are generated using Latin hypercube sampling to efficiently generate an accurate representation of the normal distribution (Stein, 1987).
Figure 4.6. Resulting coefficient of variation for the modeled concentration results when each parameter is perturbed by a normal distribution defined by a coefficient of variation of 0.1. The c and f after each parameter indicate whether it is in the course or fine grained porous media.
Figure 4.7. Coefficient of variation for the resultant concentration at each time step along the center-line of the column. The bottom of each line represents the outer edge of the course silica in the direction of gravity and the top represents the center of the fine-grained silica. Each column represents an individual parameter perturbed with a normal distribution centered at the optimized value with a coefficient of variation of 0.1. The red line shows a coefficient of variation of 0.1 for a given parameter. Each row is a different model time that corresponds to the times presented in figure 4.3.

4.4.4 Discussion

MRI is a useful tool for measuring concentration change within a medium sand column. Non-invasive imaging in three dimensions allows for the movement of concentration across a sharp material interface to be observed and modeled. Using MRI, we are able to show that small differences in density (0.3%) create flow fields that enhance the movement of concentration from small layers of low permeability material.

A density difference of 0.3% produces velocities out of the fine silica on the order of $10^{-6}$ m/s, as seen in figure 4.4. The flow moves through the fine silica at a slow speed causing the contrast agent to be transported in the direction of gravity. The speed of the circulating fluid is larger in regions where a large density gradient is present.
The model was rerun with density driven convection removed. The time required to reduce the concentration in the fine-grained material to 2.5 g/L is 14 days when density driven convection is included and 149 days when density effects are not included. It is evident that small density differences between fluids reduce the time required for mixing to occur in reservoirs where fluid movement is naturally low by creating density driven convection currents while the density of layered fluids remain in an unstable arrangement.

4.5 Acknowledgements

The authors thank Floyd Vilmont for his assistance in designing and building the sand column used for the experiments conducted in this paper. His expertise was instrumental in constructing an appropriate sand column for use in MRI.

4.6 Funding Sources

This material is based upon work supported by the National Science Foundation under Grant No. 1523488. Any opinion, findings, and conclusions or recommendations expressed in the material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

4.7 References


Faisal, T.F., S. Chevalier, and M. Sassi (2013), Experimental an numerical studies of density driven natural convection in saturated porous media with application to CO₂


Osuga, T. and S. Han (2004), Proton magnetic resonance imaging of diffusion of high- and low-molecular-weight contrast agents in opaque porous media saturated with


CHAPTER 5 Investigation of the Buildup of Concentration at the Interface between Fine and Course Soils

Contaminant transport across interfaces between fine and course soil is a difficult phenomenon to measure using traditional experimental techniques such as probes. Magnetic resonance imaging (MRI) offers the opportunity to observe transport across this interface using non-invasive measurements that quantify the concentration at high resolution. We use MRI to observe a build-up of contrast agent at the interface between fine and course silica when transport is dominated by diffusion.

5.1 Introduction

Remediation of contaminated sites is required by the clean water act and safe drinking water act and regulated by the Environmental Protection Agency. Current regulations require mitigation to be conducted until measured concentrations are below standards set by the EPA, and with an estimated 235,000 to 355,000 contaminated sites in the United States. The EPA estimates that cleanup costs through 2033 may exceed $200 billion (EPA, 2004; NRC, 2013). The cost associated with this cleanup is in large part caused by the time investment required for pump-and-treat methods, which is increased by breakthrough curve tailing caused by soil heterogeneities. This dissertation will investigate the movement of contaminants between low and high permeability materials. This movement in heterogeneous materials, specifically those with defined clay or silt layers, plays a role in contaminant tailing.

Experimentally observing fluid movement in porous media has been limited to the use of point sensors such as TDR probes, effectively limiting the spatial resolution, espe-
cially at interfaces between fine and course grained materials. For this research, the use of magnetic resonance imaging (MRI) will be employed to refine the spatial resolution to a scale of 1.5 mm x 1.5 mm x 7 mm. MRI allows for relatively large soil columns to be examined in three-dimensions for a direct comparison with numerical models with similar resolutions. This intermediate scale experiment will investigate the time required using pump-and-treat methods to remove contaminants from low permeability lenses and to calibrate a model that can then be used for field scale studies.

5.1.1 The challenges of groundwater remediation

Remediating contaminated groundwater has been a goal of the Environmental protection agency since the passage of the Safe Drinking Water Act in 1972, and the role of the Federal Government in cleanup was further enhanced by the Comprehensive Environmental Response, Compensation and Liability Act (commonly referred to as CERCLA or Superfund) in 1980. Since that time, the common method for remediation has been pump and treat. Pump and treat technology relies on extracting groundwater from contaminated sites to remove any contaminants in that region (Mackay and Cherry, 1989). The challenge arises in heterogeneous aquifers where lenses of soil with low permeabilities retard flow. These regions present a challenge by trapping and slowly releasing contaminants for many years (MacDonald and Kavanaugh, 1994). The slow release of contaminants is known as tailing, in which the concentration breakthrough curve deviates from the theoretical curve for a homogeneous region. Figure 5.1 shows the theoretical and tailing curves for contaminant removal. When tailing is present, increasing the amount of pore volumes removed (pumping time) has decreased efficiency in reducing
contaminant concentrations. This means that over the lifetime of a pump-and-treat system efficiency is lost and money is wasted due to tailing.

![Theoretical breakthrough curve and actual breakthrough curve related to the volume of fluid pumped.]

**Figure 5.1.** Theoretical breakthrough curve and actual breakthrough curve related to the volume of fluid pumped.

### 5.1.2 The effect of geologic heterogeneity on transport modeling

Transport across at the sharp interface between relatively low and high permeability porous media may not be completely described by Fickian diffusion (*Petropoulos et al.*, 2009; *Sternberg*, 2004). This creates discrepancies between experimentally observed breakthrough curves and modeled breakthrough curves both at the lab-scale and at the field-scale (*Cherblanc et al.*, 2007). The differences are described by a tailing phenomenon where the breakthrough curve is elongated when compared to the breakthrough curves of homogeneous material (*Huang et al.*, 1995).

Tailing has been investigated by using various packing of fine and course grained material. Flow from high permeability to low permeability results in more pronounced tailing than experiments with flow from low to high permeability. The differences be-
tween flow direction is further exacerbated by reduced flow speed indicating that heterogeneity plays a larger role in diffusive transport than advective transport (Berkowitz et al. 2009). Non-Fickian buildup of concentration has been observed at the interface of the fine and course grained material when flow is from course to fine (Gao, et al., 2009; Zhang and Qiao, 2010).

Cortis and Zoia (2009) used random walks with skewed Gaussians to model the impact of flow direction and the sharp material interface demonstrated by Berkowitz et al. (2009). They were able to reproduce experimental results and showed that mass was accumulating at the sharp interface. Furthermore, they showed sharp gradients at the interface as well. They hypothesize that in the transition region particles have unequal probabilities of moving into one of the distinct materials. This unequal probability creates a random force field that impacts the flow of particles across the interface.

Investigating the accumulation of mass at the interface of two materials requires that the interface be observable. This can be done using high-speed cameras and dyes (Zinn et al., 2004), or penetrating radio waves can be used to investigate porous media using methods such as x-ray (Tidwell et al., 2000) and MRI (Greiner et al., 1997).

5.1.3 Nuclear Magnetic Resonance in Porous Media

Nuclear magnetic resonance (NMR) is the absorption and re-emission of radio frequency (RF) energy by elements, most commonly hydrogen in the form of H₂O. The nuclei of elements with an odd number of protons and/or electrons have an intrinsic magnetic moment and angular momentum (spin) which is nonzero, when these elements are placed within a strong uniform magnetic field their spins align along the magnetic field
lines. Spins align in either a high-energy state (opposite to the magnetic field) or a low energy state (with the magnetic field). The distribution of alignments will favor the low energy state, such that more protons have spins aligned with the magnetic field. The nucleus can be forced into the high-energy state with RF pulses at a specific frequency – each nucleus in a given magnetic field will have a specific resonant frequency which causes it to enter the high-energy state. Nuclei will naturally revert to their low energy alignment by emitting an RF pulse that is sensitive to the material around the resonant nucleus. Through varying the magnetic field with gradient coils, areas can be targeted and sequentially scanned to construct an image of signal intensity.

NMR has many practical applications for non-invasively imaging objects, most commonly people, specifically the brain. However; NMR has also been applied to material science to study the structure of porous media. Issa and Mansfield (1994) used T1 weighted maps to estimate permeability in sandstones using a relationship of T1 of porous media and T1 of water to determine the porosity. Komlosh et al. (2011) determined the average porosity of glass capillary tubes on the voxel scale using double pulsed-field gradient MRI. In addition to mapping porosity, concentration of magnetic tracers such as Gd-diethylenetriaminepentaacetate (Gd-DTPA\(^{2-}\)) can be tracked through time. Haber-Pohlmeier et al. (2010) demonstrate that Gd-DTPA\(^{2-}\) can be tracked within a soil column using MRI to determine flow and transport behaviors on a fully 3D scale. Additionally, they determined that Gd-DTPA\(^{2-}\) is a conservative tracer due to its anionic nature and bulky structure.

Alternative contrast agents include NiCl\(_2\) (Pearl et al., 1993) and CuSO\(_4\) (Greiner et al., 1997), the requirement for a contrast agent is a paramagnetic compound that
changes the relaxation time of the solution that it is dissolved in. Greiner et al. (1997) packed two small columns with glass beads and calculated the concentration of copper sulfate at the outlet with both MRI and atomic absorption spectroscopy. The results for both homogeneous and heterogeneous bead packing showed that MRI could accurately measure outflow breakthrough curves with the added benefit of quantifying the concentration at a fine voxel resolution within the column.
5.2 Experimental Setup

The column used in this section is identical to the column used in Chapter 4. The difference is in the size of the center fine-grained silica and the care taken to eliminate convective transport. Two methods were used to eliminate convection caused by density differences between the contrast agent and water. The first method was to slowly rotate the column at a rate of one revolution every ten minutes (0.1 rpm). The second method used salt water in place of deionized water. The salt water was prepared to have an identical density to the injected contrast agent, this allowed the column to be left stationary. Figure 5.2 shows the size of the fine-grained silica in the first portion of this experiment.

![Figure 5.2. As built schematic of the plastic soil column.](image_url)
5.2.1 Column Design

The column is built from plastic without the use of metal because any metal used in the construction would prevent the columns use in the MRI scanner. Non-magnetic metal can be used safely in the scanner, but may cause distortions in the images by interrupting the radio signals used in the column. Figure 5.3 shows the individual components used to assemble the column.

![Figure 5.3. Components of the soil column.](image)

The column was assembled by gluing the two collars to either end of the clear PVC pipe. The collar has 12 threaded holes for the plastic bolts to allow the diffuser plate
to be firmly attached to the top and bottom of the column. Two of the bolts on one end had a plastic washer, figure 5.4, that allow the column to be pushed against an end plate, aligned to a marker on the end plate, and wedged in place on the MRI table. This ensures that the column is in the same position for each scan. The diffuser plate has a groove cut into the inner edge of the ring of bolts. This groove houses a 10-inch diameter O-ring that seals the column when the bolts are firmly fastened. Each diffuser plate has a ¼ inch quick connect ball valve attached to it. The two ball valves are the entry and exit points for fluid. The diffuser plate contains concentric circular grooves connected by radial spokes. Fluid enters either end of the column through the ball valve at the center of the diffuser plate and then spreads across the diffuser plate to evenly distribute flow across the entire column. Filter paper and the filter support plates are placed between the diffuser plate and the porous media in the column to prevent the porous media from entering the grooves in the diffuser plate or the tubing connected to the ball valves. The filter support is made from a perforated plastic sheet cut to fit the inner dimensions of the clear PVC pipe.
Porous media is added to the column once one end is firmly fastened. The course sand used was AFS 50-70 testing sand from US Silica and was screened with a neodymium magnet and a funnel to remove any magnetic particles from the manufacturing process that may distort the MRI signal. Metal particles were removed from the magnet and funnel after approximately 50 mL of sand was poured past the magnet. This was done until the desired amount of sand had been screened. The sand was added to water already

Figure 5.4. End of the soil column. The plastic washers and alignment marker used to align the column in the same position on the MRI table for each scan.
in the column to ensure that it was fully saturated. The sand was initially filled to the level where the bottom of the fine silica was to be located.

There are two methods that were used to place the silica core within the column. The first method was used in chapter 4 and for the rotating column and the second method was used in the column filled with salt water. The first method is to add the fine silica to a mold with water in it and freeze the saturated silica at a temperature of -20°C to prevent it from thawing too quickly. The mold should either be easy to extract the frozen silica from, such as a Shelby tube, or easy to remove from the sample, such as a disposable sauna tube. The ends of the frozen cylinder were cut to the desired length after extraction from the mold and it was placed into the center of the column, see figure 5.5. Course sand was then added to the space around and above the frozen silica.
Figure 5.5. Frozen silica placed into the center of the column.

The second method used to place the fine silica was to saturate the silica in a PVC pipe of the desired diameter without freezing it. The pipe was capped at one end during the filling process. It was then inverted and placed at the desired location. Without removing the pipe, saturated course sand was added around the pipe to the level where the fine silica ended, see figure 5.6. The PVC pipe was slowly raised to allow the course sand
to fill in the gap were the pipe had been. Any extra fine silica was removed before adding
the remaining course sand to the column.

![Figure 5.6. PVC pipe filled with fine silica surrounded by saturated course sand.](image)

Water was pumped through the filled column after both filling methods for 4
hours at a rate of 100 mL/min. This resulted in 3 pore volumes of water moving through
the column.
5.2.2 Rotating the Column

The device used to rotate the column consists of a metal frame made from plated steel slotted angle, castors for the device to rotate on, a 90V DC gearmotor for high torque and low rpm operation, and a rubber belt. Figure 5.7 shows the device in operation. The slow rotation of the column continuously reorients the direction that gravity acts on the column. The idea is to eliminate convection by continuously changing the direction in which convection would occur. Because convection within the soil column is very slow, the rotation can also be slow and thus does not introduce strong centripetal forces on the fluid. The centripetal acceleration from 0.1 rpm at the maximum radius of the column is \(1.43 \times 10^{-5} \text{ m/s}^2\).
Figure 5.7. Device used to slowly rotate the soil column.

5.2.3 Salt Water Mixture

Salt water was mixed so that the density of salt water would match the density of the injected contrast agent. This required 2.8 g/L of salt to match the 3 mmol/L of Magnevist (2.8 g/L) that was used in this experiment. The column was fully saturated with the salt water at the time of its construction.
5.3 Experimental Procedure

The saturated column was imaged prior to adding contrast agent to get baseline $T_1$ values for course sand and fine silica. After the initial scan, gadolinium based contrast agent in the form of Magnevist was pumped into the column at a rate of 100 mL/min for 1.5 hours at a concentration of 2.8 g/L or 2.6 g/L. The column was then imaged immediately after the pumping of the contrast agent stopped.

The two methods for eliminating gravity were then employed. The first method tried was the slow rotation of the column. In this case the column was stored in the MRI facility on the device shown in figure 5.7 which continuously rotated the column when it was not being scanned. This experiment took place from July of 2015 until January of 2016. The second method did not involve rotating the column and relied on using salt water with the same density as the contrast agent in place of the DI water. In this experiment, the column was stored in a consistent horizontal position in the MRI facility when not being imaged from June of 2017 until July of 2017.

Both experiments were imaged using $T_1$-weighted images. The rotating experiment was flushed with DI water at a flow rate of 100 mL/min for 1.5 hours after 3 weeks. The column continued to be imaged and flushed every 2 weeks for 6 months. The experiment with salt water was not flushed, but instead excavated after 3 weeks to quantify the concentration of contrast agent across the interface.

5.4 Results

$T_1$-weighted scans were conducted for both experiments using the same methodology presented in 4.3.2. The measured $T_1$ value depends on the concentration of Magnevist...
vist as well as the surrounding material. We found that the scaling factor of 4.1 found by Tofts (2010) is accurate for converting $T_1$ values to concentration when Magnevist is in pure water or the course sand. However, when the Magnevist is in the fine-grained silica the concentration calculated from the $T_1$ values requires a different scaling factor. We found this scaling factor by creating seven samples with concentrations of Magnevist ranging from 0 to 3 mmol/L. The value of $1000/T_1$ in the test tubes with fine silica is plotted against the known concentration values in mmol/L in the test tubes with water in them. Figure 5.8 shows the resulting trend lines for the tubes with water and the tubes with silica, the upper trend line for the silica indicates that a scaling factor of 5.3 results in the same concentrations as the comparable tubes with water in them using the known scaling factor of 4.1.

![Figure 5.8](image_url)

**Figure 5.8.** $T_1$ to concentration scaling factor found using varying concentrations of Magnevist in pure water and test tubes containing fine silica.

### 5.4.1 Rotating Column

The rotating column experiment was conducted from July of 2015 to January of 2016. The initial three-week diffusion into the fine-grained silica is shown in figure 5.9.
The Magnevist was injected from the top of the figure, the tail that is shown is caused by a region where the velocity is zero as the flow lines converge around the fine-grained silica.

The distortions and noise around the edges of the column, such as the area at the lower left of each panel of figure 5.9 is caused by the edges of the column being at the maximum field of view for the MRI. The scanner has a drop off in both the transmit radio frequency and the receive field. Luckily, the area that we are concerned with is in the center of the field of view.

From panel A to D of figure 5.9, we see the development of high concentration outline of the fine-grained soil. This outline is more apparent when a line scan is taken across each image. Figure 5.10 shows the matching line scan taken by vertically averaging over 10 voxels along the horizontal centerline of the column from left to right. Panels B, C, and D of figure 5.10 all show an anomalous decrease and increase in the concentration of Magnevist at the interface between the fine-grained silica and the course sand.
Figure 5.9. Sagittal plane concentration scan of the rotated column over a three-week period. The concentration of Magnevist is shown in g/L. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection.
Figure 5.10. Horizontal line scan of 10 vertically averaged voxels taken across each panel of figure 5.9. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection.

5.4.2 Salt Water Column

The rotational experiment was reproduced using salt water instead of DI water so that the density of the Magnevist could be matched. The experiment was conducted from June to July of 2017. Again, the initial three-week diffusion into the fine-grained silica is
shown in figure 5.11. The radius of the fine-grained silica was adjusted so that the column did not have to be frozen in a sauna tube. The signal at the edges of the field of view was weaker in this experiment and consequently the T₁ values could not be fit due to the noise that this created.

Two things immediately stand out in this image, the first is that the tail is skinnier and the second is that the high concentration outline of the fine-grained silica noticed in figure 5.9 is present in this experiment as well. The smaller tail is caused by the fine-grained silica cylinder having a smaller radius, this means that there is a smaller region where the velocity is zero because the fluid doesn’t take as long of a path around the fine-grained cylinder. The increased concentration at the interface between the two materials is more pronounced in this second experiment, and it appears to be thicker than in figure 5.9.

The line scan images in figure 5.12 of the salt water experiment shows that the increased concentration is indeed thicker at the interface. The increased concentration was confirmed using inductively coupled plasma mass spectrometry ICP-MS. Samples were taken from six points within the column, shown in figure 5.13. Each sample contained approximately 5 mL of soil. Samples were diluted with 10 mL of trace metal grade nitric acid diluted to 5% and vigorously shaken to ensure that all gadolinium within the sample was properly dissolved. The samples were then centrifuged at 8,000 rpm for 10 minutes and the supernatant was extracted and sent to Meadowlands Environmental Research Institute for ICP-MS analysis for Gd. The results of the ICP-MS are presented in table 5.1.
Figure 5.11. Sagittal plane concentration scan of the column with salt water over a three week period. The concentration of Magnevist is shown in g/L. A) 0 days after initial injection of Magnevist. B) 9 days after initial injection. C) 17 days after initial injection. D) 21 days after initial injection.
Figure 5.12. Horizontal line scan of 10 vertically averaged voxels taken across each panel of figure 5.11. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection.
Figure 5.13. ICP-MS sample locations from central cross section of the soil column from figure 5.11 and 5.12.
Table 5.1. ICP-MS results for 6 samples shown in figure 5.13. The Gd to Magnevist factor is used to account for the fact that the MRI is measuring concentration of Magnevist while ICP-MS is measuring the concentration of only the Gadolinium which makes up 16.8% of the molecular weight of Magnevist.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Conc. [ppb]</th>
<th>Conc. RSD %</th>
<th>Dilution Factor</th>
<th>Conc. [mg/L]</th>
<th>Dilution Factor</th>
<th>Gd to Magnevist factor</th>
<th>Conc. [g/L]</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Top Edge</td>
<td>58.521</td>
<td>1.658</td>
<td>1000</td>
<td>58.521</td>
<td>6.66</td>
<td>5.96</td>
<td>2.323</td>
</tr>
<tr>
<td>#2 Center</td>
<td>26.624</td>
<td>0.579</td>
<td>1000</td>
<td>26.624</td>
<td>6.66</td>
<td>5.96</td>
<td>1.057</td>
</tr>
<tr>
<td>#3 Lower Center</td>
<td>39.516</td>
<td>1.585</td>
<td>1000</td>
<td>39.516</td>
<td>6.66</td>
<td>5.96</td>
<td>1.569</td>
</tr>
<tr>
<td>#4 Bottom Inner Edge</td>
<td>56.436</td>
<td>0.281</td>
<td>1000</td>
<td>56.436</td>
<td>6.66</td>
<td>5.96</td>
<td>2.240</td>
</tr>
<tr>
<td>#5 Bottom Edge</td>
<td>60.554</td>
<td>0.760</td>
<td>1000</td>
<td>60.554</td>
<td>6.66</td>
<td>5.96</td>
<td>2.404</td>
</tr>
<tr>
<td>#6 Bottom Outer Edge</td>
<td>59.292</td>
<td>0.766</td>
<td>1000</td>
<td>59.292</td>
<td>6.66</td>
<td>5.96</td>
<td>2.354</td>
</tr>
</tbody>
</table>

The results from ICP-MS indicate that the buildup of concentration at the interface of the fine silica and course sand, samples 1 and 5, as measured by the MRI are real when compared to samples 4 and 6. The ICP-MS results are plotted against the MRI data, taken 4 days earlier than the samples were extracted, in figure 5.14.
Figure 5.14. ICP-MS concentration of Magnevist compared to MRI measured Magnevist concentrations 4 days prior.

The buildup of concentration at the interface occurs in both experiments in which convection due to a difference in the fluid densities is removed. The likely cause of this buildup is due to greater levels of adsorption taking place in the fine silica because of a higher surface area. When concentration initially diffuses into the fine silica, Magnevist adsorbs onto the silica near the interface before mobile Magnevist is able to penetrate further. This creates a region where the total concentration, in both the fluid phase and solid phase, is higher than that in the surrounding soil regions. Buildup occurs when the porous media that the contaminant is diffusing into has a higher adsorption capacity than the porous media it is diffusing from.
5.4.3 Model Results

The MRI results were compared to the numerical model developed in Chapter 4. Removing gravitational convection from the model allows the mesh to be reduced to a 2d axial-symmetric case where only one slice of the column needs to be simulated. The model was able to fairly accurately reproduce the observed results including the concentration buildup at the interface by adjusting parameters for the hydraulic conductivity of the fine silica, the molecular diffusion for both porous media, and the maximum adsorption coefficient of the fine silica. Figure 5.15 shows the comparison of the model to the observed data.

![Graphs showing concentration over time](image)

Figure 5.15. Observed MRI results of the saltwater column, shown as blue circles, are compared to the modeled total concentration in black, the fluid phase concentration in magenta, and the solid phase concentration in green. A) 0 days after initial injection of Magnevist. B) 4 days after initial injection. C) 11 days after initial injection. D) 20 days after initial injection.
Table 5.2. Model parameters. Values correspond to the Course Sand and Fine Sand. Highlighted values indicate parameters that differ from those presented in table 4.1.

<table>
<thead>
<tr>
<th>Sand Layer</th>
<th>Permeability $k$ (m$^2$)</th>
<th>Transverse Dispersivity $\alpha_T$ (m)</th>
<th>Longitudinal Dispersivity $\alpha_L$ (m)</th>
<th>Molecular Diffusion $D_m$ (m$^2$/s)</th>
<th>Langmuir Coefficient $K_{La}$ (m$^3$/kg)</th>
<th>Max. Adsorption $b$ (kg/m$^3$)</th>
<th>Adsorption Speed $a_a$ (s$^{-1}$)</th>
<th>Desorption Speed $a_d$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>6.78e-11</td>
<td>2.5e-5</td>
<td>2.5e-4</td>
<td>4.0e-10</td>
<td>1.98</td>
<td>8.22</td>
<td>6.06e-12</td>
<td>1.28e-11</td>
</tr>
<tr>
<td>Fine</td>
<td>1.00e-14</td>
<td>2e-6</td>
<td>2e-5</td>
<td>4.5e-10</td>
<td>7.16</td>
<td>1.4</td>
<td>3.31e-6</td>
<td>5.39e-10</td>
</tr>
</tbody>
</table>

The model is able to reproduce the observed buildup because the Magnevist adsorbs to the the fine silica at the edge of the column as it diffuses into the cylinder before it is able to continue moving into the column. This delays the time it takes for the Magnevist to reach the center of the column and causes the total concentration, the concentration adsorbed to the silica plus the concentration in solution, to be greater than the concentration in the course sand. The difference generated by the model is larger than that observed by either MRI images or from ICP-MS.

### 5.4.4 Mesh Convergence

The model was run using multiple mesh densities to ensure convergence. The true concentration values were taken to be the modeled concentration on a very fine mesh. For this problem, the mesh resolution was 0.833 mm in the z-axis and 0.625 mm in the x axis, the y-axis is set as a single wedge of the column. The results of six successive mesh refinements is shown in figure 5.16. The mesh converges at a y-resolution of around 50 volumes.
Figure 5.16. Infinity norm of the modeled concentration with successive mesh refinements. The y-resolution refers to the number of volumes in the y-axis. The ratio of the y- and z-axis was kept constant with each refinement.

5.5 Discussion

MRI is able to quantify concentration in three dimensions and allows for experiments which target the interaction at the interface to be observed. The apparent buildup at the interface was shown using MRI when density driven convection was eliminated by either rotating the column or by using saltwater with the same density as the contrast agent. The likely explanation for the observed buildup is explained using the OpenFOAM model. The buildup likely occurs because the Magnevist adsorbs to the fine silica at the edge of the cylinder before fluid can continue diffusing towards the center. This delays the time it takes for the Magnevist to reach the center of the column and causes the total concentration, the concentration adsorbed to the silica plus the concentration in solution, to be greater than the surrounding concentration in the course sand. Other possible expla-
nations that need to be looked at in more depth include electrostatic forces and differential diffusion across the boundary (Alvarez-Ramirez et al., 2014). Electrostatic forces could play a role when Magnevist is used as a contrast. Magnevist is an ionic compound that dissolves into three particles, Gd-DTPA$^{2-}$ and two particles of meglumine$^+$ (Laurent et al., 2006). Negatively charged silica particles in the form of the porous media used may form electrostatic bonds with the meglumine. Differential diffusion may occur because of a sharp interface between both porosity and diffusion coefficients, on the microscale, this may cause Gd-DTPA particles to interact with more fine silica particles. Obeying the principle of Occam’s razor means that the simplest model is most likely correct. In this case the adsorption model can fairly accurately reproduce the experimental results. However, the concentration difference at the interface generated by the model is larger than that observed by either MRI images or from ICP-MS.
CHAPTER 6 Conclusions

This work has explored the potential for pathway development from formations targeted by hydraulic fracturing to shallower formations and the atmosphere. Two methods were used to determine whether the formation of a path was possible. The first method looked at pathways that are formed from the induced fractures from a new large-scale, horizontal well intersecting an adjacent preexisting well. The second method used a large dataset of oil and gas wells constructed in Alberta Canada to identify characteristics that lead to pathway formation in existing wells. Additionally, this work has used magnetic resonance imaging (MRI) to quantify the concentration of contrast agent in a soil column. Multiple experiments were run to isolate convection and diffusion from each other in the transport equation.

The first half of this work was designed to address the risk that hydraulic fracturing poses to water resources and the environment. I showed that, with enough data, hazard maps can be created for the formation of a pathway by means of an adjacent well. This requires known locations and depths of a large percentage of wells as well as the induced fracture growth from enough fracture jobs to develop an appropriate fracture growth distribution. Hazard maps for the area of New York underlain by the Marcellus shale were estimated for a range of typical hydraulic fracturing operations\(^1\); and thus, a lot of uncertainty could be eliminated on a well by well basis when new wells are proposed since the size, location, and target depth would be known prior to construction. The probabilities of encountering a pathway as defined in this work ranged from 0.0% in are-

\(^{1}\) A hydraulic fracturing operation is defined as the number of horizontal wells emanating from the same surface pad. The number of wells could range anywhere from 1 to as many as 12 horizontal wells (Manda et al., 2014).
as that did not have any existing wells at sufficient depth to encounter the induced fractures to as high as 56.8% when a large enough fracturing operation was placed in areas with a history of oil and gas development. Figure 6.1 shows a representation of the range of properties tested. A hydraulic fracturing (HF) area of influence of 1,500 ft$^2$ represents a very small fracturing operation while a 1,600,000 ft$^2$ HF area of influence represents a fracturing job that is on the larger end, but not out of the range of possibility. This larger estimate assumes that fractures are complex and create a permeable zone around the horizontal section of the well that extends 400 ft. in either direction from 8 wells. Each well is fractured along 1/8 of its 4,000 ft. horizontal length.
Figure 6.1. Percent probability of encountering an existing well with a new horizontal hydraulic fracturing job in New York State underlain by the Marcellus Shale. There are 60 wells in each calculation point and the HF area of influence is $1,500 \text{ ft}^2$, $30,000 \text{ ft}^2$ and $1,600,000 \text{ ft}^2$, respectively. Contours are delineated at the color bar labels and are in log scale.
The second portion of assessing pathway formation involved collecting data on well construction, production, and locations from Alberta Canada. Wells in the region used in this study have been tested for gas migration and surface casing vent flow since 1995. This data set was skewed having only 12.65% of the dataset classified as providing a pathway. Additionally, many wells were missing data points and required either predictive methods that could deal with missing data or a method to impute missing values. In this work, the missing values were imputed data using an iterative principal component analysis (PCA) fitting method called trimmed scores regression (TSR) (Artega and Fererrer, 2002).

Seven predictive algorithms were compared to random guessing to determine which method performed the best on this data set. A random forest was determined to produce the best results for this data set when compared against backpropagation networks, support vector machines, and generalized regression. Random forests correctly identify 73% of the wells in the validation data set and had the highest geometric mean between sensitivity and specificity of the models tested at 68%. The features that were determined to be the most influential as to whether a well would provide a pathway were the age of the well, whether or not it deviated from vertical, and the casing size.

The second half of this work used MRI to investigate flow and transport in porous media. MRI was shown to be an effective tool in quantifying the concentration of the paramagnetic contrast agent Magnevist in porous media. Transport under the influence of density dependent convection was shown to occur with a density difference as small as 0.03% and was successfully modeled using OpenFOAM. Transport was then investigated when density driven convection was eliminated, first by slowly rotating the column and
then by using salt water at an identical density to the contrast agent. Both methods appear
to have eliminated the effect of density driven convection. When convection was elimi-
nated from the experiment, a buildup of concentration was observed at the interface be-
tween a fine ground silica core and a courser surrounding sand layer. The buildup could
be reproduced using OpenFOAM, but had a larger peak than was observed using either
MRI or ICP-MS to quantify the concentration.
Comprehensive Bibliography


Alpha Environmental Consultants Inc. (2009), Issues related to developing the Marcellus Shale and other low-permeability gas reservoirs, Prepared for NYSERDA, Albany, NY.


BC Oil & Gas Commission (2010), Communication during fracture stimulation, BC Oil & Gas Commission, Victoria, British Columbia.


Fitzgibbon, J.J. (1984), Sintered spherical pellets containing clay as a major component useful for gas and oil well proppants, U.S. Patent No. 4,427,068.


Ground Water Protection Council (2009), State oil and natural gas regulations designed to protect water resources, Prepared for The U.S. Department of Energy/National Energy Technology Laboratory, Washington, D.C.


Hall, L.D. (1964), Nuclear magnetic resonance, Advances in Carbohydrate Chemistry, 19, 51-93.


Osborn, S.G., A. Vengosh, N.R. Warner, and R. B. Jackson (2011), Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing,


Stein, M. (1987), Large sample properties of simulations using latin hypercube sampling, Technometrics, 29 (2), 143-151.


Tofts, P.S. (2010), T1-weighted DCE imaging concepts: Modelling, Acquisition and Analysis, MAGNETOM Flash, 30-39.


U.S. EPA (2016), Hydraulic fracturing for oil and gas: Impacts from the hydraulic fractureing water cycle on drinking water resources in the United States, U.S. Environmental Protection Agency, Washington, DC.


Chapter 2 of this dissertation was run on the Vermont Advanced Computer Cluster (VACC) using distributed computing so that many different parameters could be tested in an efficient manner. The data for wells was preprocessed and stored as a MatLab data file. Well data included the location of the wells, the depth of the wells, the depth to the base of the target shale formation, and a shapefile outline for the target shale formation. The formation outline is needed to determine whether existing wells fall within the target formation. This allows a statewide database to be used with multiple formations within the same state.

The following MatLab code is responsible for calculating the probability of encountering an existing well at each representative area (RA). RAs are calculated such that there is one area every mile in the x and y direction.

```matlab
% Probability of encounter for NY underlain by Marcellus shale
function [] = prob_run(num_wells,padarea)

%addpath('m:/Papers/Fracking_Probability_Model/Matlab_Model/igmm-master/');
load('Marcellus_wells.mat');

%% distribute points throughout the Marcellus Shale, use a uniform % distribution of 1 evaluation point per square mile
num_wells = str2num(num_wells);
padarea = str2num(padarea);
name = [num2str(padarea),'_mi2_',num2str(num_wells),'_'];

% discretize region into representative areas based on lat/lon % conversion to miles
area = 1;
xdisc = sqrt(area)*5280/267522.23;
ydisc = sqrt(area)*5280/364476.65;
```
xx = min(M_xy(:,1)):xdisc:max(M_xy(:,1));
yy = min(M_xy(:,2)):ydisc:max(M_xy(:,2));
[xg, yg] = meshgrid(xx,yy);
temp = zeros(size(xg,1),size(xg,2));

% find all points that are within the Marcellus Shale
for i=1:length(S)
    yv = [lat{i},lat{i}(1)];
xv = [lon{i},lon{i}(1)];
    IN = inpolygon(xg,yg,xv,yv);
temp(IN) = 1;
end

% remove those points that are outside of the Marcellus
xg(temp==0) = [];
yg(temp==0) = [];
xg = xg';
yg = yg';

%% for each evaluation point find the nearest x points, defined by num_wells

% set up matrices
M_xy(M_depth==0,:) = [];
M_depth(M_depth == 0) = [];
for run = 1:10
    prob = zeros(size(xg,1),1);
aprob = zeros(size(xg,1),1);
    fpdist_mu = 5.979; % predetermined mean for fracture propagation
fpdist_sig = 0.579; % predetermined standard deviation prop.
    centarea = zeros(size(xg,1),1);
    numfit = 20; % number of Gaussian Mixture Models to try to fit

% for each representative area
for i = 1:size(xg,1)
    % distance from the RA to each well
    dist = deg2sm(distance([M_xy(:,2),M_xy(:,1)],yg(i),xg(i)));
    [val,ind] = sort(dist);
    % take the num_wells nearest wells to the RA
    eval_depths = M_depth(ind(1:num_wells));
    % area of the circle that captures the closest num_wells
    centarea(i) = ... 
        pi.*((dist(ind(num_wells+1))+dist(ind(num_wells)))/2).^2;

    % remove zero depths from data due to incomplete data
    eval_depths(eval_depths==0) = [];

    % develop distributions for depth data
temp = eval_depths;

    % fitting criteria for Gaussian Mixture Model
    bic = zeros(1,numfit);
aic = zeros(1,numfit);
% gmdistribution options
options = statset('MaxIter',500);

% try each fit from 1 Gaussian until numfit Gaussians
for j=1:numfit
    % sometimes gmdistribtion.fit will encounter an error
    try
        obj = gmdistribution.fit(temp,j,'Options',options);
        wdepth(j).mu = obj.mu';
        wdepth(j).pi = obj.PComponents';
        wdepth(j).s = ...
            1./reshape(obj.Sigma(1,1,:),1,size(obj.Sigma,3));
        bic(j) = obj.BIC;
        aic(j) = obj.AIC;
    catch
        aic(j:numfit) = inf;
        bic(j:numfit) = inf;
        break;
    end
end

% the best fit is the value with the minimum akaike information
% criterion
[val bestfit] = min(aic);
wdepth = wdepth(bestfit);

% errors occur if there is a single spike that has a variance
% smaller the discretization of our pdf, this is due to the
% integration using trapz which is just the trapezoid method over
% our discretization. If this occurs we need to ensure that the
% variance is increased slightly to extend to the edges of our
% discretization.
wdepth.s(wdepth.s>1) = 1;

% Determine the depth of the shale in the RA
shale_depth = mean(M_sd(ind(1:num_wells)));

% set up a matrix containing the probability of a well being at a
% given depth in the RA based on the fit Gaussian Mixture Model and
% the depth of the shale in the area.
dx = 1;
xs = 0:dx:max(shale_depth,max(temp))*1.1;
well_dist = zeros(size(xs,1),size(xs,2));
for j=1:length(wdepth.mu)
    well_dist = well_dist + ...
        wdepth.pi(j)/sqrt(2*pi*1./wdepth.s(j)) * ...
        exp(-(xs-wdepth.mu(j)).^2/(2*1./wdepth.s(j)));
end

% find the fracture distribution using a log-normal pdf
fpdist = lognpdf(xs,fpdist_mu,fpdist_sig);

% find the shale depth distribution, for this work, a single spike
% is put at the depth of the shale representing the shale depth
shaledist = zeros(size(xs,1),size(xs,2));
shaledist(round(shale_depth)) = 1;

% convolve the well depth with the fracture propagation
total_dist = conv(well_dist,fpdist).*dx;

% size of the matrix increases as a result of convolution
nx = 0:dx:xs(end)*2;

% Cross-correlation of the shale depth and the result of the
% previous calculation
total_dist = xcorr(shaledist,total_dist).*dx;

% size increases again
nx = sort([-nx,nx(2:end)]);

% integrate the values that have a resultant depth less than 0
prob(i) = trapz(total_dist(nx<=0));

% Find the horizontal probability of encountering a well
if centarea(i) < padarea
    aprob(i) = 1;
else
    aprob(i) = (1-(1-padarea/centarea(i))^num_wells);
end
end

% Multiple the vertical and horizontal probabilities to get the total
% probability
tprob = aprob.*prob;

% Save the results
save([{'prob_'},name,num2str(run),'','mat'],{'tprob','centarea'})
end

Results are saved in the directory in which the program was run under the names
prob_<padarea>_mi2_<num_wells>_<run>.mat. An example being:
prob0.00107_mi2_60_1.mat.

A.1 Running the code on the VACC

Running the code is done through a script from the user node on the VACC.
There are two scripts that are needed to run the code. The first sets up the paths and exe-
cutes the MatLab file on the VACC node that the job is assigned to. The second submits the job to the queue.

A.1.1 run_prob_run.sh

#!/bin/sh
# script for execution of deployed applications
# Sets up the MCR environment for the current $ARCH and executes
# the specified command.
# exe_name=$0
exe_dir=`dirname "$0"`
echo "------------------------------------------"
if [ "x$1" = "x" ]; then
echo Usage:
echo $0 \<deployedMCRroot\> args
else
echo Setting up environment variables
MCRROOT="$1"
echo ---
LD_LIBRARY_PATH=.:${MCRROOT}/runtime/glnxa64 ;
LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${MCRROOT}/bin/glnxa64 ;
LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${MCRROOT}/sys/os/glnxa64;
   MCRJRE=${MCRROOT}/sys/java/jre/glnxa64/jre/lib/amd64 ;
   LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${MCRJRE}/native_threads ;
   LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${MCRJRE}/server ;
   LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${MCRJRE}/client ;
   LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:${MCRJRE} ;
XAPPLRESDIR=${MCRROOT}/X11/app-defaults ;
export LD_LIBRARY_PATH;
export XAPPLRESDIR;
echo LD_LIBRARY_PATH is ${LD_LIBRARY_PATH};
shift 1
"${exe_dir}"/prob_run $*
fi
exit

A.1.2 Marcellus.script

# PBS -l nodes=1:ppn=1
# PBS -l walltime=30:00:00
# PBS -M jmontagu@uvm.edu
# PBS -m bea
export LD_LIBRARY_PATH =
$LD_LIBRARY_PATH:/gpfs1/arch/x86_64/matlab2010a/bin/glnxa64:
/gpfs1/arch/x86_64/matlab2010a/runtime/glnxa64:
/gpfs1/arch/x86_64/matlab2010a/sys/os/glnxa64:
A.1.3 Submitting a large array of jobs with different variables

Submitting multiple jobs to the VACC requires input into the terminal from the directory where the MatLab files are located. This is where the parameters for the number of wells in each RA and the horizontal area that each pad takes up is defined as nw and pa respectively. Note that pad areas are defined in square miles.

```bash
../run_prob_run.sh /gpfs1/arch/x86_64/matlab2010a/mcr/v713 $nw $pa
```

```bash
>for i in "${pareas[@]}"
>do
>   for j in "${nwells[@]}"
>      do
>         export pa="$i"
>         export nw="$j"
>         echo $pa
>         echo $nw
>         qsub -v pa,nw Marcellus.script -l walltime=30:00:00 -l nodes=1:ppn=1,pmem=12gb,pvmem=13gb
>      done
>done
```
APPENDIX B Wellbore Pathway Prediction Code

The prediction of wellbore pathways was done using a number of predictive algorithms. Each of the sections of code used to produce the results of that work is presented below as MatLab code. This work makes use of many of the built-in MatLab predictive algorithms and requires the Neural Network toolbox. This work relies heavily on preprocessing the dataset to ensure that it is formatted properly. That code is also reproduced below.

B.1 Preprocessing

%% Jim Montague
% Alberta Test Area Data normalizing and formatting to be ready for use
% in neural networks and other data classification methods

clear all;close all;clc;

% add path (Windows)
addpath('M:/Alberta_Well_Analysis/...Test_Area_Well_Classifications/Matlab_Files')

% Load data
emptys = xlsread('Alberta_Test_Area_Data_W_Density_Emptys.xlsx','A3:AR38394');

% data names
CPName = {'Longitude','Latitude','depth(m)','op ind','N_Op_wells',...
'Op_Size','const_year','aban_year','years_abandoned','dev (deg)',...
'prod_hrs','cum_fluid','case_size','lower_depth (m)','cons_time',...
'Period_Blank','Period_Cambrian','Period_Devonian',...
'Period_Lower_Cretaceous','Period_Lower_Devonian',...
'Period_Middle_Cambrian','Period_Middle_Devonian',...
'Period_Mississippian','Period_Ordovician','Period_Precambrian',...
'Period_Tertiary','Period_Upper_Cambrian',...
'Period_Upper_Cretaceous','Period_Upper_Devonian','Type_Blank',...
'Type_Carbonates','Type_Clastics','Type_Evaporites','Type_Shale',...
'GHV','Vapor_Pressure','Molecular_wt','Rel_Density',...
'Critical_Pressure','Critical Temp','density_(w/sqmi)',...
'Fail_class','Fail_type','Fail/Not');

% load imputed values
load('Full_impute.mat') % built from MDIgui command which requires
% MDI toolbox 4 available from: http://mseg.webs.upv.es/Software_e.html
%% Normalize the data and format for use with data classifiers
fail_class = wellData(:,42);

% make sure to exclude the classifications that depict the type of % failure
Y = [wellData(:,3)';wellData(:,5)';wellData(:,7)';...
    wellData(:,9:15)';wellData(:,35:41)'];

t = [fail_class==-1,fail_class==0,fail_class==1]';
Rt = t;
Rt(3,Rt(2,:)==1) = 1; % classify minor as failed
% Rt(1,Rt(2,:)==1) = 1; % classify minor as not failed
Rt = [Rt(1,:);Rt(3,:)];

% fill each column's nan's with the imputed values
RW = MDIToolbox_results.X_imputed';

% Normalize RW between 0 and 1 replacing nan's with the mean of the % k-nearest neighbors, there shouldn't be any nan's when using the % imputed values
RW = knearest(RW,10);

% take a random subset as a training and validation set because the % full data set is skewed heavily towards wells that have not % failed, the training set should also be skewed, but not as heavily or % the algorithms will not work well
% set a random number generator to produce reproducible samples
rng(1)
GW = RW(:,Rt(1,:)==1);
FW = RW(:,Rt(2,:)==1);
samples = min(size(GW,2),size(FW,2));%694;
trainsize = min(round(samples*2/3),600);
coef=1.25; % ratio between failed and non-failed wells
pGW = randperm(size(GW,2),size(GW,2));
pFW = randperm(size(FW,2),size(FW,2));
trainRW = [GW(:,pGW(1:round(trainsize*coef))),FW(:,pFW(1:trainsize))];
trainRt = [ones(1,round(trainsize*coef)),zeros(1,trainsize);zeros(1,round(trainsize*coef)),ones(1,trainsize)];
valRW = [GW(:,pGW(trainsize+1:end)),FW(:,pFW(trainsize+1:end))];
valRt = [ones(1,samples-trainsize),zeros(1,samples-trainsize);zeros(1,samples-trainsize),ones(1,samples-trainsize)];

% The data sets are now:
% RW and Rt for the full normalized set
% trainRW and trainRt for the training set
% valRW and valRt for the validation set (which contains all remaining % data not used for training)
save('Alberta_Test_Area_Normalized_Data.mat','RW','Rt','trainRW',...
    'trainRt','valRW','valRt');

% write all to csv files
csvwrite('RW.csv',RW);
% make a separate group of data that only contains complete sets
fail_class = emptys(:,42);
emptyY = [emptys(:,3)';emptys(:,5)';emptys(:,7)';emptys(:,9:15)';emptys(:,35:41)'
emptyT = [fail_class==-1,fail_class==0 | fail_class ==1
TrainSet = emptyY(:,sum(isnan(emptyY),1)==0);
TrainAns = emptyT(:,sum(isnan(emptyY),1)==0);

comp_dataRW = TrainSet;
comp_dataRt = TrainAns;

FullDataRW = RW;
FullDataRt = Rt;

% normalize the data
comp_dataRW = knearest(comp_dataRW,MDItoolbox_results.X_imputed');

rng(1)
GW = FullDataRW(:,FullDataRt(1,:)==1);
FW = FullDataRW(:,FullDataRt(2,:)==1);
samples = min(size(GW,2),size(FW,2));
trainsize = min(round(samples*2/3),4000);  % ratio between failed and non-failed wells
coef=1.25;  % ratio between failed and non-failed wells
pGW = randperm(size(GW,2),size(GW,2));
pFW = randperm(size(FW,2),size(FW,2));

FDtrainRW = ...
[GW(:,pGW(1:round(trainsize*coef))),FW(:,pFW(1:trainsize))];
FDtrainRt = [ones(1,round(trainsize*coef)),zeros(1,1:trainsize);
zeros(1,round(trainsize*coef)),ones(1,1:trainsize)];

% validation size
GWvalsize = size(GW(:,round(trainsize*coef)+1:end), 2);
FWvalsize = size(FW(:,trainsize+1:end),2);
FDvalRW = [GW(:,pGW(round(trainsize*coef)+1:end)),...
FW(:,pFW(trainsize+1:end))];
FDvalRt = [ones(1,GWvalsize),zeros(1,FWvalsize);
zeros(1,GWvalsize),ones(1,FWvalsize)];

% All Data
FullDataRW = [GW(:,pGW(round(trainsize*coef)+1:end)),...
FW(:,pFW(trainsize+1:end))];
numValG = length(pGW(round(trainsize*coef)+1:end));
numValF = length(pFW(trainsize+1:end));
FullDataRt = [ones(1,numValG),zeros(1,numValF);
zeros(1,numValG),ones(1,numValF)];

save('compData.mat','comp_dataRW','comp_dataRt');
save('FullData.mat','FullDataRW','FullDataRt',...
    'FDtrainRW','FDtrainRt','FDvalRW','FDvalRt');

B.2 Random Guess

%% Random selection for n trials
clc;clear all;close all;
load('Alberta_Test_Area_Normalized_Data.mat');
load('FullData.mat');
n = 30; % number of trials

%% Make matrices to hold values for each trial
train_results = zeros(2,2,n);
fd_results = zeros(2,2,n);
all_results = zeros(2,2,n);

%% Find percentage of each class
train_percent_fail = sum(FDtrainRt(2,:))/size(FDtrainRt,2);
fd_percent_fail = sum(FullDataRt(2,:))/size(FullDataRt,2);
all_percent_fail = sum(Rt(2,:))/size(Rt,2);

%% iterate through n trials to find the random guess values
for i = 1:n
    rng(i)
    % train comparison
    train_rand = rand(1,size(FDtrainRt,2))>train_percent_fail;
    train_results(:,:,i) = confusion_mat(train_rand,FDtrainRt(1,:));

    % FD comparison
    fd_rand = rand(1,size(FullDataRt,2))>fd_percent_fail;
    fd_results(:,:,i) = confusion_mat(fd_rand,FullDataRt(1,:));

    % all comparison
    all_rand = rand(1,size(Rt,2))>all_percent_fail;
    all_results(:,:,i) = confusion_mat(all_rand,Rt(1,:));
end
save('RG.mat','train_results','fd_results','all_results');

B.3 Back Propagation

% Alberta Test Area BackProp network
clc;clear all;close all;
% load Alberta test area normalized data
load('Alberta_Test_Area_Normalized_Data.mat');
load('FullData.mat');
n = 30;
%% make matrices to hold values for each trial
train_results = zeros(2,2,n);
fd_results = zeros(2,2,n);
all_results = zeros(2,2,n);
classes_train = zeros(n,size(FDtrainRt,2));
classes_test = zeros(n,size(FullDataRt,2));
classes_all = zeros(n,size(Rt,2));

%% set up neural net with a hidden layer of size 20

% Bayesian Regularization
for i = 1:n
    rng(i) % use a different but consistent random seed for each trial
    net = patternnet(20,'trainFcn','trainbr');

    % train data and set up the network
    [net,tr] = train(net,FDtrainRW,FDtrainRt);
    cutoff = .5;
    % For Diagnostic Purposes
    % figure
    % plotconfusion(Rt,testAll);

    % Training Results
    disp('train')
    trainset = net(FDtrainRW);
    trainset(1,:) = trainset(1,:)>=cutoff;
    trainset(2,:) = trainset(2,:)>(1-cutoff);
    classes_train(i,:) = trainset(1,:);
    train_results(:,:,i) = confusion_mat(trainset(1,:),FDtrainRt(1,:));

    % Validation Results
    disp('test')
    trainset = net(FullDataRW);
    trainset(1,:) = trainset(1,:)>=cutoff;
    trainset(2,:) = trainset(2,:)>(1-cutoff);
    classes_test(i,:) = trainset(1,:);
    fd_results(:,:,i) = confusion_mat(trainset(1,:),FullDataRt(1,:));

    % All Results
    disp('All')
    testAll = net(RW);
    testAll(1,:) = testAll(1,:)>=cutoff;
    testAll(2,:) = testAll(2,:)>(1-cutoff);
    classes_all(i,:) = testAll(1,:);
    all_results(:,:,i) = confusion_mat(testAll(1,:),Rt(1,:));

    % save the trained model in case we want to come back to these
    % results without training again
    BR_Model{i} = net;
end

% save the results of classification
save('BR.mat','train_results','fd_results','all_results',
     'classes_train','classes_test','classes_all','BR_Model');

%% set up neural net with a hidden layer of size 20
% Gradient Descent
for i = 1:n
    rng(i)
    net = patternnet(20,'trainFcn','traingdx');
    
    % train data
    [net,tr] = train(net,FDtrainRW,FDtrainRt);
    
    cutoff = .5;
    % figure
    % plotconfusion(Rt,testAll);
    
    disp('train')
    trainset = net(FDtrainRW);
    trainset(1,:) = trainset(1,:)>cutoff;
    trainset(2,:) = trainset(2,:)>1-cutoff;
    classes_train(i,:) = trainset(1,:);
    
    train_results(:,:,i) = confusion_mat(trainset(1,:),FDtrainRt(1,:));
    train_results(:,:,i)
    
    disp('test')
    trainset = net(FullDataRW);
    trainset(1,:) = trainset(1,:)>cutoff;
    trainset(2,:) = trainset(2,:)>1-cutoff;
    classes_test(i,:) = trainset(1,:);
    
    fd_results(:,:,i) = confusion_mat(trainset(1,:),FullDataRt(1,:));
    fd_results(:,:,i)
    
    disp('All')
    testAll = net(RW);
    testAll(1,:) = testAll(1,:)>cutoff;
    testAll(2,:) = testAll(2,:)>1-cutoff;
    
    classes_all(i,:) = testAll(1,:);
    all_results(:,:,i) = confusion_mat(testAll(1,:),Rt(1,:));
    
    GDX_Model{i} = net;
end
save('GDX.mat','train_results','fd_results','all_results',
     'classes_train','classes_test','classes_all','GDX_Model');

%% set up neural net with a hidden layer of size 20
% Levenberg-Marquardt
for i = 1:n
    rng(i)
    i
net = patternnet(20,'trainFcn','trainlm');

% train data
[net,tr] = train(net,FDtrainRW,FDtrainRt);

cutoff = .5;
% figure
% plotconfusion(Rt,testAll);

disp('train')
trainset = net(FDtrainRW);
trainset(1,:) = trainset(1,:)>=cutoff;
trainset(2,:) = trainset(2,:)>(1-cutoff);

classes_train(i,:) = trainset(1,:);
train_results(:,:,i) = confusion_mat(trainset(1,:),FDtrainRt(1,:));

train_results(:,:,i)

disp('test')
trainset = net(FullDataRW);
trainset(1,:) = trainset(1,:)>=cutoff;
trainset(2,:) = trainset(2,:)>(1-cutoff);

classes_test(i,:) = trainset(1,:);
fd_results(:,:,i) = confusion_mat(trainset(1,:),FullDataRt(1,:));

disp('All')
testAll = net(RW);
testAll(1,:) = testAll(1,:)>=cutoff;
testAll(2,:) = testAll(2,:)>(1-cutoff);

classes_all(i,:) = testAll(1,:);
all_results(:,:,i) = confusion_mat(testAll(1,:),Rt(1,:));

LM_Model{i} = net;
end
save('LM.mat','train_results','fd_results','all_results','classes_train','classes_test','classes_all','LM_Model');

% Random Forest

clc;clear all;close all;
load('Alberta_Test_Area_Normalized_Data.mat');
load('FullData.mat');
n = 30;
try
    load('compData.mat');
    comp = true;
    disp('complete data present')
catch
    comp = false;
%% make matrices to hold values for each trial
train_results = zeros(2,2,n);
fd_results = zeros(2,2,n);
all_results = zeros(2,2,n);
if comp
    if size(comp_dataRW,1) > size(RW,1)
        comp_dataRW(11:16,:) = [];
    end
    comp_results = zeros(2,2,n);
    classes_comp = zeros(n,size(comp_dataRt,2));
end

classes_train = zeros(n,size(FDtrainRt,2));
classes_test = zeros(n,size(FullDataRt,2));
classes_all = zeros(n,size(Rt,2));
importanceOfFeature = zeros(n,size(FDtrainRW,1));

for i = 1:n
    rng(i)
    i

    minleaf = 1; % minimum leaf size
    numtree = 200; % size of the tree

    % train using MatLabs TreeBagger method
    treebag = TreeBagger(numtree,FDtrainRW',FDtrainRt(1,:)',...
        'OOBPredictorImportance','On','Method','classification');

    classes_train(i,:) = str2num(cell2mat(predict(treebag,FDtrainRW')));
    train_results(:,:,i) = confusion_mat(classes_train(i,:),...
        FDtrainRt(1,:));
    train_results(:,:,i)

    % test
    classes_test(i,:) = str2num(cell2mat(predict(treebag,FullDataRW')));
    fd_results(:,:,i) = confusion_mat(classes_test(i,:),...
        FullDataRt(1,:));
    fd_results(:,:,i)

    % all
    classes_all(i,:) = str2num(cell2mat(predict(treebag,RW')));
    all_results(:,:,i) = confusion_mat(classes_all(i,:),Rt(1,:));
    all_results(:,:,i)

    importanceOfFeature(i,:) = treebag.OOBPermutedVarDeltaError;

    if comp
        classes_comp(i,:) = str2num(cell2mat(predict(treebag,...
            comp_dataRW')));
        comp_results(:,:,i) = confusion_mat(classes_comp(i,:),...
            comp_dataRt(1,:));
    end
end
B.5 Support Vector Machine

% Support vector machine
clc; clear all; close all;
load('Alberta_Test_Area_Normalized_Data.mat');
load('FullData.mat');

% n = 30;
n = 1; % is not affected by rng

%% make matrices to hold values for each trial
train_results = zeros(2,2,n);
fd_results = zeros(2,2,n);
all_results = zeros(2,2,n);

classes_train = zeros(n,size(FDtrainRt,2));
classes_test = zeros(n,size(FullDataRt,2));
classes_all = zeros(n,size(Rt,2));

% support vector machine parameters
kern_scale = 1;
box_constraint = 1;

for i = 1:n
    rng(i)
    SVMIn = FDtrainRW';
    SVMOut = FDtrainRt(1,:)';

    SVMModel = fitcsvm(SVMIn,SVMOut,'KernelFunction','rbf', ...
    'KernelScale',kern_scale,'BoxConstraint',box_constraint, ...
    'Standardize',true);
    classes_train(i,:) = predict(SVMModel,SVMIn);
    CVMdl1 = crossval(SVMModel);
    misclass1 = kfoldLoss(CVMdl1);
    disp('train')
    train_results(:,:,i) = confusion_mat(classes_train,FDtrainRt(1,:));
    train_results(:,:,i)
% Find optimal scale and constraint %%%%%%%%%%%%%%%%%
c = cvpartition(size(SVMOut,1),'KFold',10);
minfn = @(z1)kfoldLoss(fitcsvm(SVMIn,SVMOut,'CVPartition',c,...
'KernelFunction','rbf','BoxConstraint',exp(z2),...  
'KernelScale',exp(z1),'Standardize',true));
minfn = @(z1)kfoldLoss(crossval(fitcsvm(SVMIn,SVMOut,...
'KernelFunction','rbf','BoxConstraint',exp(z2),...
'KernelScale',exp(z1),'Standardize',true)));
 opts = optimset('TolX',5e-4,'TolFun',5e-4);
m = 20;
fval = zeros(m,1);  
z = zeros(m,2);  
for j = 1:m;  
    [searchmin fval(j)] = fminsearch(minfn,randn(2,1),opts);  
    z(j,:) = exp(searchmin);  
end  
z = z(fval == min(fval),:)
% End optimize %%%%%%%%%%%%%%%%%

B.6 Generalized Regression Neural Network

% General Regression Neural Network for Well Data
clc;clear all;close all;
load('Alberta_Test_Area_Normalized_Data.mat');
load('FullData.mat');
rng(1); % not affected by randomization
n = 1;

% make matrices to hold values for each trial
train_results = zeros(2,2,n);

% General Regression Neural Network for Well Data

\textbf{train\_raw} = zeros(2,4,n);
\textbf{fd\_results} = zeros(2,2,n);
\textbf{fd\_raw} = \textbf{train\_raw};
\textbf{all\_results} = zeros(2,2,n);
\textbf{all\_raw} = \textbf{train\_raw};

\textbf{classes\_train} = zeros(n,\text{size(FDtrainRt,2)});
\textbf{classes\_test} = zeros(n,\text{size(FullDataRt,2)});
\textbf{classes\_all} = zeros(n,\text{size(Rt,2)});

\textbf{y} = FDtrainRt(1,:);
\textbf{x} = FDtrainRW;

\textbf{dimo} = \text{size(y,1)}; % output dimension
\textbf{dimi} = \text{size(x,1)}; % input dimension

% weights
\textbf{wp} = \textbf{x}; % set to input parameters
% set ws to the outputs
\textbf{ws} = zeros(\text{size(y,2)},\text{size(y,1)+1});
\textbf{ws}(:,\text{size(y,1)}) = \textbf{y};
\textbf{ws}(:,\text{end}) = \text{ones(size(ws,1),1)}; % the last value of ws is set to one

\textbf{sig} = [0.09]; % fit parameter

\textbf{x1} = FullDataRW;
\textbf{c1r} = zeros(\text{size(sig,2)},1);
\textbf{c1w} = \textbf{c1r};
\textbf{c2r} = \textbf{c1r};
\textbf{c2w} = \textbf{c1r};
\textbf{acc} = \textbf{c1r};
\textbf{precision} = \textbf{c1r};
\textbf{NPV} = \textbf{c1r};
\textbf{NMI} = zeros(\text{size(sig,2)},1);

\textbf{for} i = 1:n
\textbf{for} k = 1:\text{size(sig,2)}
% Train
\textbf{x1} = FDtrainRW;
\textbf{y1} = GRNN\_fit(\textbf{x1},\textbf{wp},\textbf{ws},\text{sig(k)},\text{dimo});

\textbf{interp\_results} = \textbf{y1};
\textbf{interp\_results(\textbf{interp\_results} \geq .5)} = 1;
\textbf{interp\_results(\textbf{interp\_results} < .5)} = 0;
\textbf{classes\_train(i,:) = interp\_results};
[\textbf{train\_results,:,:}, \textbf{train\_raw,:,:}] = ...\text{confusion\_mat(y1,FDtrainRt(1,:)})

% Verification
\textbf{x1} = FullDataRW;
\textbf{y1} = GRNN\_fit(\textbf{x1},\textbf{wp},\textbf{ws},\text{sig(k)},\text{dimo});

% act\_results = FullDataRt(1,:);%y;%'FullDataRt(1,:);
\textbf{interp\_results} = \textbf{y1};
\textbf{interp\_results(\textbf{interp\_results} \geq .5)} = 1;
\textbf{interp\_results(\textbf{interp\_results} < .5)} = 0;
classes_test(i,:) = interp_results;
[fd_results(:,:,i), fd_raw(:,:,i)] = ...
    confusion_mat(y1,FullDataRt(1,:))

% All Data
x1 = RW;
y1 = GRNN_fit(x1,wp,ws,sig(k),dimo);

interp_results = y1;
interp_results(interp_results >=.5) = 1;
interp_results(interp_results < .5) = 0;
classes_all(i,:) = interp_results;
[all_results(:,:,i), all_raw(:,:,i)] = confusion_mat(y1,Rt(1,:))
end

end
save('GRNN.mat','train_results','fd_results', ...
'all_results','train_raw','fd_raw','all_raw','classes_train', ...
'classes_test','classes_all');

B.6.1 GRNN Fit

function [y] = GRNN_fit(x,wp,ws,sig,dimo)
% Interpolates the GRNN function based on the weights wp and ws and the
% parameter sig. The output is the interpolated values at every x loca-
% tion. From (Donald Specht, 1991)

y = zeros(dimo,size(x,2));
for j=1:size(x,2)
    Ij = zeros(1,size(wp,2));
    for k=1:size(wp,2);
        Ij(1,k) = sum((wp(:,k)-x(:,j)).^2);
    end
    FI = exp(-Ij./(2*sig^2));
    A = FI*ws(:,1:dimo);
    B = FI*ws(:,end);
    y(j) = A./B;
end
end

B.7 Generalized Regression

% linear regression
clc;clear all;close all;

load('Alberta_Test_Area_Normalized_Data.mat');
load('FullData.mat');
n = 1; % not controlled by randomization

% Make matrices to hold values for each trial
train_results = zeros(2,2,n);
fd_results = zeros(2,2,n);
all_results = zeros(2,2,n);

classes_train = zeros(n,size(FDtrainRt,2));
classes_test = zeros(n,size(FullDataRt,2));
classes_all = zeros(n,size(Rt,2));

%% train models
cat = FDtrainRt';

% conduct n random trials
for j = 1:n
    rng(j);
    mdl = fitglm(FDtrainRW',cat(:,1),'Distribution','binomial');
    coeff = mdl.Coefficients{:,1};

    % Train
    raw_results = zeros(size(FDtrainRW,2),1);
    for i =1:size(FDtrainRW,2)
        raw_results(i) = coeff(1) + coeff(2:end)'*FDtrainRW(:,i);
    end

    act_results = FDtrainRt(1,:);'
    interp_results = raw_results;
    interp_results(interp_results >=.5) = 1;
    interp_results(interp_results < .5) = 0;
    classes_train(j,:) = interp_results;
    train_results(:,:,j) = confusion_mat(interp_results,FDtrainRt(1,:))

% Test
Test = FullDataRW;
act_results = FullDataRt(1,:);'
raw_results = zeros(size(Test,2),1);
for i =1:size(Test,2)
    raw_results(i) = coeff(1) + coeff(2:end)'*Test(:,i);
end

interp_results = raw_results;
interp_results(interp_results >=.5) = 1;
interp_results(interp_results < .5) = 0;
classes_test(j,:) = interp_results;
fd_results(:,:,j) = confusion_mat(interp_results,FullDataRt(1,:));

% Full Data
Test = RW;
act_results = Rt(1,:);'
raw_results = zeros(size(Test,2),1);
for i =1:size(Test,2)
    raw_results(i) = coeff(1) + coeff(2:end)'*Test(:,i);
end

interp_results = raw_results;
interp_results(interp_results >=.5) = 1;
interp_results(interp_results < .5) = 0;
classes_all(:, :) = interp_results;
all_results(:, :, j) = confusion_mat(interp_results, Rt(1, :));

GLR_Model = mdl;
end

save('GLR.mat', 'train_results', 'fd_results', 'all_results', 'classes_train', 'classes_test', 'classes_all', 'GLR_Model');

**B.8 Confusion Matrix**

```matlab
function [confusion, raw_conf] = confusion_mat(model, actual)
% take in model results and predicted results and return the confusion
% matrix. Model and actual are each 1 x n where n is the number of
% samples. Each sample is classified as an integer value corresponding
% to its class. The output, confusion, is therefore a c x c matrix,
% where c is the number of unique integer values in actual.

% preprocess inputs to reshape if they are the wrong orientation or
% throw an error if dimensions are incorrect.
if size(model, 1) > 1
    if size(model, 2) > 1
        error('Model dimensions incorrect, must be 1 x n, where n... is the number of samples');
    end
    model = model';
end

if size(actual, 1) > 1
    if size(actual, 2) > 1
        error('Actual dimensions incorrect, must be 1 x n, where n... is the number of samples');
    end
    actual = actual';
end

u_thresh = .66;
l_thresh = .33;

int_model = model;
int_model(int_model >= .5) = 1;
int_model(int_model < .5) = 0;
n_model = model;
n_model(n_model >= u_thresh) = 1;
n_model(n_model <= l_thresh) = 0;
% used for uncertain class in raw confusion matrix
n_model(n_model < .5 & n_model > l_thresh) = .25;
n_model(n_model < u_thresh & n_model > .5) = .75;
subclass = [.25 .75];

classes = unique(actual); % all possible classifications from data
c = size(classes, 2); % determines the size of the confusion matrix
```

158
confusion = zeros(c);
raw_conf = zeros(c,c+2); % adds an uncertain class

for i = 1:c
    for j = 1:c
        confusion(i,j) = sum(actual == classes(i) & ...
                         int_model == classes(j));
        raw_conf(i,j) = sum(actual == classes(i) & ...
                         n_model == classes(j));
    end
    % adds in the uncertain classes
    for j = c+1:c+2
        raw_conf(i,j) = sum(actual == classes(i) & ...
                         n_model == subclass(j-2));
    end
end

B.9 Accuracy Measure

function [acc_meas] = accuracy_measure(confusion_mat)
% analyze results from convo network take in a confusion matrix
% (only works for 2x2 at the moment. Returns a data structure of
% accuracy measures
a = confusion_mat(1,1);
b = confusion_mat(1,2);
c = confusion_mat(2,1);
d = confusion_mat(2,2);

% Equations for assessing the accuracy of a classification algorithm
%
% + -
% + a b
% - c d

x = [a b;
     c d];

N = sum(sum(x));
NMI = 0;
xip = sum(x,2);
xpj = sum(x,1);
for i = 1:size(x,1)
    for j = 1:size(x,2)
        NMI = NMI + 1/N*x(i,j)*log(x(i,j)*N/(xip(i)*xpj(j))));
    end
end
NMIb = 0;
for j = 1:size(x,2)
    NMIb = NMIb - (xpj(j)/N*log(xpj(j)/N));
End
% accuracy
acc = sum(diag(x))/N;
[a b;c d]
disp(['Accuracy: ',num2str(acc)])

% sensitivity
sens = a/(a+b);
disp(['sensitivity: ',num2str(sens)])

% specificity
spec = d/(c+d);
disp(['specificity: ',num2str(spec)])

if acc<=.5
    s = -1;
else
    s = 1;
end

% check
% nmi2 = 1-(-a*log(a)-b*log(b)-c*log(c)- ...
%   d*log(d)+(a+b)*log(a+b)+(c+d)*log(c+d))/(N*log(N)-...
%   ((a+c)*log(a+c)+(b+d)*log(b+d)))

% Normalized mutual information
NMI = s*NMI/NMIb;
NMI = (NMI+1)./2; % between 0 and 1

disp(['NMI: ',num2str(NMI)])

% Proportion of specific agreement
psa = 2*a/(2*a+b+c);
disp(['proportion of specific agreement: ',num2str(psa)])

% cohen's kappa
% random both correct
rbc = (a/(a+b)*d/(d+c));

% random both wrong
rbw = (b/(a+b)*c/(d+c));

% probability of random agreement
pra = rbc*rbw;
kappa = (acc-pra)/pra;
disp(['kappa: ',num2str(kappa)])

% sPPAa
sPPAa = 1/2*(a/(a+c)+a/(a+b));
disp(['sPPAa: ',num2str(sPPAa)])

% sPPAp
sPPAp = (a/(a+c)*a/(a+b));
disp(['sPPAp: ',num2str(sPPAp)])

% geometric-mean
gmean = sqrt(sens*spec);
disp(['gmean: ',num2str(gmean)])

% Matthews correlation coefficient
MCC = ((a*d)-(c*b))/sqrt((a+c)*(a+b)*(d+c)*(d+b));
MCC = (MCC+1)./2; % between 0 and 1

% add metrics to data structure
acc_meas.accuracy = acc;
acc_meas.NMI = NMI;
acc_meas.psa = psa;
acc_meas.sPPAa = sPPAa;
acc_meas.gmean = gmean;
acc_meas.MCC = MCC;
acc_meas.sensitivity = sens;
acc_meas.specificity = spec;
OpenFOAM comes in three varieties, which stem from various breakpoints in the original development and acquisition of the software. Figure C.1 shows the history and development of OpenFOAM from its original creation in 2004. The three main OpenFOAM distributions are the Foam-Extended project driven by the FOAM community, OpenFOAM maintained by the OpenFOAM foundation, and OpenFOAM + maintained by ESI. Each version is slightly different, so care must be taken when reading online guides as to which version the guide is referring to and which version you are working with. The version used for the purposes of this dissertation is OpenFOAM 1606+ available at openfoam.com.

Figure C.1. OpenFOAM development history and breakpoints.
C.1 Installation of OpenFOAM +

Note: You must be running a linux operating system to be able to modify the source files of OpenFOAM.

• Go to openfoam.com
• Select “Download” from the top ribbon
• Click on “Current release”
• Select “Source code”
• Download the 2 tarballs, OpenFOAM-vxxxx+.tgz and ThirdParty-vxxxx+.tgz
• Follow the instructions to unpack and build OpenFOAM found at:
  openfoam.com/download/install-source.php
  and
  openfoam.com/code/build-guide.php
• Make sure that all prerequisites are installed prior to compiling OpenFOAM.
• It is recommended to build many of the third party files from their respective folders in your newly created ThirdParty folder.

C.2 Modifying OpenFOAM

Once OpenFOAM has been compiled and tested, there are several additions that need to be made to run the transport codes used in this dissertation. The codes are reproduced below and available at https://github.com/jamesmontague12/OpenFOAM_Transport. Navigate to your OpenFOAM folder then open applications, solvers, incompressible and replace the pisoFoam folder with the one found at the above github address. Then open up GWpisoFoam and from the terminal enter wmake to build the GWpisoFoam solver. Next, add
MRI_3d_Gravity to the base level of you OpenFOAM installation /OpenFOAM/. You should be able to run the case used in this dissertation from the terminal using the command ./Allrun in the terminal.

C.3 Optimization

Optimization is achieved in this dissertation using pySOT. The pySOT setup code is available at the above github link in the MRI_3d_Gravity folder. To run the optimization you must first install pySOT. This can be done by following the instructions at https://pysot.readthedocs.io/en/latest/. It is recommended that you read the installation instructions carefully so you get all of the required packages installed.

Optimization can then be run directly from MRI_3d_Gravity by entering python pySOT_Opt.py. The optimization process takes a very long time, so do not expect immediate results. Further work could be done to speed up the optimization by taking advantage of the parallel processing functionality that is built into pySOT.

C.4 Modifying the OpenFOAM case files

The case file is modified primarily by manipulating the files found in the System folder. These files are: controlDict, fvSchemes, fvSolution, setFieldsDict, snappyHexMeshDict, and topoSetDict. Each of these files is an OpenFOAM dictionary that provides runtime information at the start of each model run. The following sections will go through each of these dictionaries.

C.4.1 controlDict
The file controlDict contains information for how long to run the simulation, how large of time steps to take, how often to output results, and where and what to sample (this is used for optimization). The following is an example of the controlDict used for the work in this dissertation. All time units are given in seconds and locations are given in meters. The controlDict file used for this dissertation is provided below. Additional comments have been added to the code presented below to describe what each section does in greater detail.

```plaintext
// The following is the standard heading used in all OpenFOAM files and // must be present. 
/*--------------------------------*- C++ -*-------- ------------------*/
| \      /  F ield         | OpenFOAM: The Open Source CFD Toolbox |
| \    /   O peration     | Version:  2.3.0                      |
|   \  /    A nd           | Web:      www.OpenFOAM.org            |
|    \/     M anipulation  |                                         |
\*------------------------------------------------- ------------------*/
FoamFile
{
    version 2.0;
    format ascii;
    class dictionary; // Line 12 defines the type of file as a
    // dictionary file
    location "system"; // folder where the file is located
    object controlDict; // name of the file, this must match
    // exactly, note that there is no file
    // extension.

    application GWpisoFoam; // The application that we want run
    startFrom startTime; // the starting time for this simulation
    startTime 0; // the time that the simulation will start at
    stopAt endTime; // the ending time
    endTime 4320000; //50 days Jan 28 - Mar 19
    deltaT 900; //15 min // the time step that is taken
    writeControl timeStep;
    writeInterval 96; // (daily) number of time steps between each output
    purgeWrite 0; // when to purge output files
    writeFormat ascii;
    writePrecision 9; // how many numbers to use in outputs
    writeCompression off;
    timeFormat general;
    timePrecision 9; // how many numbers to use for time
```

165
/* The lines for the writePrecision and timePrecision (lines 28 and 31) are important to ensure that information is lost when rounding. This is especially true for time precision when dealing with long simulations. If the precision is smaller than the amount of time that the simulation is run for then the time will be expressed in truncated scientific notation. For example, if you want to express 1 day (86,400 seconds) using a timePrecision of 2, the result will be truncated to $8.6 \times 10^4$ meaning that any increment of less than 1000 seconds is lost. */

functions
{
    Probes // point samples that are output to a separate location
    {
        type          probes;
        functionObjectLibs ("libsampling.so");
        outputControl outputTime;
        probeLocations // (x y z) locations of probes
        {
            (0 0 0)
            (0 -0.0026 0)
            (0 -0.0052 0)
            (0 -0.0078 0)
            (0 -0.0104 0)
            (0 -0.013 0)
            (0 -0.0156 0)
            (0 -0.0182 0)
            (0 -0.0208 0)
            (0 -0.0234 0)
            (0 -0.026 0)
            (0 -0.0286 0)
            (0 -0.0312 0)
            (0 -0.0338 0)
            (0 -0.0364 0)
            (0 -0.039 0)
            (0 -0.0416 0)
            (0 -0.0442 0)
            (0 -0.0468 0)
            (0 -0.0494 0)
            (0 -0.052 0)
            (0 -0.0546 0)
            (0 -0.0572 0)
            (0 -0.0598 0)
            (0 -0.0624 0)
            (0 -0.065 0)
            (0 -0.0676 0)
            (0 -0.0702 0)
            (0 -0.0728 0)
            (0 -0.0754 0)
            (0 -0.078 0)
        }
    }
}
C.4.2  fvSchemes

The file fvSchemes contains information on which method the finite volume code should use for each portion of the numerical equations. The default method used is provided after the word default for each scheme. If a different method is desired for a particular equation than it can be specified after default. In this example, there are multiple divergence schemes being applied. The default scheme is to use the Gauss linear method, but two additional schemes have been applied for fluid convection and concentration convection. In both cases an upwind scheme has been employed. A list of schemes and a more
comprehensive explanation can be found at https://cfd.direct/openfoam/user-guide/fvschemes/.

FoamFile
{
    version 2.0;
    format ascii;
    class dictionary;
    location "system";
    object fvSchemes;
}

// * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * //

ddtSchemes
{
    default Euler;
}

gradSchemes
{
    default Gauss linear;
}

divSchemes
{
    default Gauss linear;
    div(phi,U) Gauss upwind;
    div(phi,C) bounded Gauss upwind;
}

laplacianSchemes
{
    default Gauss linear corrected;
}

interpolationSchemes
{
    default linear;
}

snGradSchemes
{
    default corrected;
}
The file *fvSolution* includes information on which equation solvers to use for each solved parameter. It is not recommended to change the information in this file. An explanation of the solvers can be found at [https://cfd.direct/openfoam/user-guide/fvsolution/](https://cfd.direct/openfoam/user-guide/fvsolution/).

```plaintext
FoamFile
{
    version     2.0;
    format      ascii;
    class       dictionary;
    location    "system";
    object      fvSolution;
}
// * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
//
solvers
{
    p
    {
        solver         GAMG;
        tolerance      1e-06;
        relTol         0.1;
        smoother       GaussSeidel;
        nPreSweeps     0;
        nPostSweeps    2;
        cacheAgglomeration on;
        agglomerator   faceAreaPair;
    }
```
nCellsInCoarsestLevel 10;
mergeLevels 1;

"(U|k|epsilon|omega|R|nuTilda)"
{
  solver smoothSolver;
  smoother GaussSeidel;
  tolerance 1e-05;
  relTol 0;
}

C
{
  solver BICCG;
  preconditioner DILU;
  tolerance 1e-07;
  relTol 0;
}

Cim
{
  solver smoothSolver;
  smoother GaussSeidel;
  nSweeps 2;
  tolerance 1e-07;
  relTol 0.1;
}

Cip
{
  solver smoothSolver;
  smoother GaussSeidel;
  nSweeps 2;
  tolerance 1e-07;
  relTol 0.1;
}

U
{
  solver smoothSolver;
  smoother GaussSeidel;
  nSweeps 2;
  tolerance 1e-07;
  relTol 0.1;
}

e
{
  solver smoothSolver;
  smoother symGaussSeidel;
  tolerance 1e-06;
  relTol 0.1;
}
"(k|\epsilon)"
{
  \$U;
  tolerance 1e-07;
  relTol 0.1;
}
}

PISO
{
  nCorrectors 2;
  nNonOrthogonalCorrectors 0;
  pRefCell 0;
  pRefValue 0;
}

//*******************************************************************//

C.4.4 setFieldsDict

The file setFieldsDict sets different values for the regions defined. This is extremely useful for spatially varying the values of a field at runtime start. The default-FieldValues sets the field values for all areas not defined as a region. Specific areas that have different values can be set up in the regions section. First the type of shape must be defined, for this work a cylinder is defined in the center with a radius of 0.0365125 m and a height of 0.1233 m. Each parameter must be referenced using its field type, ie volScalarFieldValue. Parameters that are not defined as volume fields cannot be modified in this manner.

FoilFile
{
  version 2.0;
  format ascii;
  class dictionary;
}
location "system";
object setFieldsDict;

// * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
//
defaultFieldValues
(
  volScalarFieldValue alphaT 2.5e-5
  volScalarFieldValue alphaL 2.5e-4
  volScalarFieldValue Rf 1
  volScalarFieldValue C 0
  volScalarFieldValue am 1e-16
  volScalarFieldValue aim 1
  volScalarFieldValue vcrit 1e-10
  volScalarFieldValue diam 100e-6
  volScalarFieldValue Dm 7.35777755312e-12
  volScalarFieldValue Klang 1.97896115885
  volScalarFieldValue beta 8.21600664978
  volScalarFieldValue aa 6.05755330852e-12
  volScalarFieldValue ad 1.28443661901e-11
  volScalarFieldValue vcrit 1e-10
  volScalarFieldValue diam 100e-6
  volScalarFieldValue Dm 7.35777755312e-12
  volScalarFieldValue Klang 1.97896115885
  volScalarFieldValue beta 8.21600664978
  volScalarFieldValue aa 6.05755330852e-12
  volScalarFieldValue ad 1.28443661901e-11
);

regions
(
  // fine sand
cylinderToCell
  {
    p1 (0 0 0.0616666667);
    p2 (0 0 -0.0616666667);
    radius 0.0365125;
    fieldValues
      {
        volScalarFieldValue alphaT 2e-6
        volScalarFieldValue alphaL 2e-5
        volScalarFieldValue am 1e-3
        volScalarFieldValue aim 5
        volScalarFieldValue C 2.09425420198
        volScalarFieldValue Cim 0.805745798019
        volScalarFieldValue Ctot 2.9
        volScalarFieldValue diam 7e-6
        volScalarFieldValue n 0.2
        volScalarFieldValue rhoS 2837
        volScalarFieldValue Dm 1.58501112013e-11
        volScalarFieldValue Klang 7.16394000712
        volScalarFieldValue beta 0.859451044429
        volScalarFieldValue aa 3.3172809501e-06
        volScalarFieldValue ad 5.38920714245e-10
        volScalarFieldValue Km 1
      }
  }
);
C.4.5 snappyHexMeshDict

The snappyHexMeshDict controls how the original rectangular mesh is fit to the desired shape. The shape is defined by a .stl file. Blender, an open source 3D object creator, was used to generate Open_Cyl.stl. SnappyHexMeshDict snaps the mesh to the object defined in geometry. Additional comments have been added below to clarify some aspects of the file. Please note that snappyHexMeshDict can create regions along the boundaries that are not smooth if care is not taken.

```
FoamFile
{
    version     2.0;
    format      ascii;
    class       dictionary;
    object      snappyHexMeshDict;
}

// each of the steps to run
castellatedMesh true;  // refines the mesh along a given boundary
snap            true;  // snapping to a surface
addLayers       true;  // adds additional layers to a surface

geometry // surface defined by an .stl file in /constant/triSurface
{
    Open_Cyl.stl
    {
        type triSurfaceMesh;
        name wall;
    }
```
castellatedMeshControls
{
    features ( );
    // refinement along each surface, refinement is done by splitting cells
    refinementSurfaces
    {
        wall
        {
            level (0 0);
            patchInfo
            {
                type wall;
            }
        }
    }
    // refinement in a specific region
    refinementRegions
    {
        locationInMesh (0 0 0);
        maxLocalCells 1000000;
        maxGlobalCells 2000000;
        minRefinementCells 0;
        nCellsBetweenLevels 1;
        resolveFeatureAngle 80.0;
        allowFreeStandingZoneFaces true;
    }
    // snapping to surface
    snapControls
    {
        nSmoothPatch 8; // how many iterations of smoothening to use
        tolerance 1.0; // relative distance maximum edge length to snap to
        // surface
        nSolveIter 300; // number of mesh displacement iterations
        nRelaxIter 5; // number of relaxation iterations
        nFeatureSnapIter 10;
        implicitFeatureSnap false;
        explicitFeatureSnap true;
        multiRegionFeatureSnap false;
    }
    // adds additional cells along a given surface
    addLayersControls
    {

layers
{
    "wall.*"
    {
        nSurfaceLayers 10; // adds 10 layers of cells perpendicular to
        // the edge
    }
}

relativeSizes false; // use relative size for thickness or absolute
// size, relative size is based on the outside
// layers thickness
    expansionRatio 1; // how much larger to expand each successive layer
// thickness
    thickness 0.050; // size of the outer cell layer
    minThickness 0.00002; // minimum thickness of a given cell
    nGrow 0;
    featureAngle 80;
    slipFeatureAngle 180;
    nRelaxIter 10;
    nSmoothSurfaceNormals 1;
    nSmoothNormals 3;
    nSmoothThickness 10;
    maxFaceThicknessRatio 0.9;
    maxThicknessToMedialRatio .8;
    minMedianAxisAngle 80;
    nBufferCellsNoExtrude 2;
    nLayerIter 200;
    nRelaxedIter 30;
}

// controls the quality of the resultant mesh
meshQualityControls
{
    maxNonOrtho 65;
    maxBoundarySkewness 20;
    maxInternalSkewness 4;
    maxConcave 80;
    minVol 1e-13;
    minTetQuality 1e-30;
    minArea -1;
    minTwist 0.05;
    minDeterminant 0.001;
    minFaceWeight 0.05;
    minVolRatio 0.01;
    minTriangleTwist -1;
    nSmoothScale 4;
    errorReduction 0.75;
    relaxed
}
maxNonOrtho 75;
}
}
dump 0;
mergeTolerance 1E-6;
// **************************************************************************/

C.4.6 topoSetDict

The topoSetDict file defines cellZones within the model that can be used as discrete areas to change parameters similar to setFieldsDict. The file used for this dissertation has two regions the fineSand and the courseSand. The fine sand is defined as a cylinder in the center of the column and the course sand is defined as all cells that aren’t already in the fineSand cellSet.

FoamFile
{
  version     2.0;
  format      ascii;
  class       dictionary;
  object      topoSetDict;
}
// * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * //

actions
{
  // fine sand
  {
    name   fineSand;
    type   cellSet;
    action new;
    source cylinderToCell;
    sourceInfo
    {
    }
p1 (0 0 0.06166666667);
p2 (0 0 -0.06166666667);
radius .0365125;// 1 cell -> .09592;
}
}

name fineSand;
type cellZoneSet;
action new;
source setToCellZone;
sourceInfo
{
    set fineSand;
}

// course sand is all the other cells
{
    name courseSand;
type cellZoneSet;
action clear;
}
{
    name courseSand;
type cellSet;
action add;
source cellToCell;
sourceInfo
{
    set fineSand;
}
}
{
    name courseSand;
type cellSet;
action invert;
}
{
    name courseSand;
type cellZoneSet;
action new;
source setToCellZone;
sourceInfo
{
    set courseSand;
}
}
);

// ********************************************************************************
C.4.7  fvOptions

fvOptions is located in the /constant folder. The file is used in this case to define an explicit porosity source to add porous resistance to the velocity and pressure equations. The file defines two porosity zones that are set in topoSetDict as cellZones. The type of porosity zone used is the Darcy-Forchheimer equation adds an inertial term which is defined by $f$ in this file. The parameter $d$ which is given in units of $1/m^2$ is the porosity divided by the permeability and is defined as a vector which allows anisotropy to be included for the permeability.

```plaintext
FoamFile
{
    version     2.0;
    format      ascii;
    class       dictionary;
    location    "constant";
    object      fvOptions;
}

porosity1
{
    type         explicitPorositySource;
    active       yes;

    explicitPorositySourceCoeffs
    {
        selectionMode cellZone;
        cellZone      fineSand;
        type          DarcyForchheimer;
        DarcyForchheimerCoeffs
        {
            d d [0 -2 0 0 0 0] (160257158084.0 160257158084.0 160257158084.0 0)
            f f [0 -1 0 0 0 0] (0 0 0);
        }
    }
}
```

178
coordinateSystem
{
    type cartesian;
    origin (0 0 0);
    coordinateRotation
    {
        type axesRotation;
        e1 (0 0 1);
        e2 (0 1 0);
    }
}

porosity2
{
    type explicitPorositySource;
    active yes;

    explicitPorositySourceCoeffs
    {
        selectionMode cellZone;
        cellZone courseSand;
        type DarcyForchheimer;
        DarcyForchheimerCoeffs
        {
            d d [0 -2 0 0 0 0] (3693196363.61 3693196363.61 3693196363.61);
            f f [0 -1 0 0 0 0] (0 0 0);
            coordinateSystem
            {
                type cartesian;
                origin (0 0 0);
                coordinateRotation
                {
                    type axesRotation;
                    e1 (0 0 1);
                    e2 (0 1 0);
                }
            }
        }
    }
}