U.S. Department of Energy National Energy Technology Laboratory

DOE Award No. DE-FE0031625

FINAL REPORT

Robust Carbon Dioxide Plume Imaging Using Joint Tomographic Inversion of Seismic Onset Time and Distributed Pressure and Temperature Measurements

Principal Investigator: Dr. Akhil Datta-Gupta

datta-gupta@tamu.edu 979-847-9030

March 25, 2022

Recipient: Texas A&M Engineering Experiment Station

College Station, TX 77845

DUNS Number: 847205572

Project Grant Period: September 1, 2018 and ending December 31, 2021

Reporting Period End Date: December 31, 2021

Report Term or Frequency: Final

Signature of Submitting Official:

Acknowledgment:
This material is based upon work supported by the Department of Energy National Energy Technology Laboratory under Award Number(s) DE-FE0031625.

TABLE OF CONTENTS

Executive Summary

We develop and demonstrate rapid and cost-effective methodologies for spatio-temporal tracking of CO2 plumes during geologic sequestration using joint inversion of seismic data and distributed pressure and temperature measurements. Key elements of our methodology are: (a) a computationally efficient approach to pressure and temperature propagation, (b) analysis of time lapse seismic data using a novel 'seismic onset time' approach to detect fluid front propagation, and (c) data assimilation and uncertainty assessment via joint inversion of pressure, temperature and time lapse seismic data, and (d) validating the numerical tomographic inversion using a CO2 injection demonstration projects, specifically data collected from the from the Petra Nova Parish Holdings CCUS project in the West Ranch Field, Texas and the Chester-16 reef CO2 injection site in Northern Michigan which is part of the DOE Midwestern Carbon Sequestration Project. The research team is led by Texas A&M University and includes Battelle as a subcontractor with support from Shell, Anadarko, Chevron and JX Nippon.

A carbon dioxide (CO2) water-alternating-gas (WAG) pilot was conducted to gain insights into tertiary oil recovery potential via CO2 flood in the West Ranch Field as part of the Petra Nova project, the world's largest post-combustion CO2 capture and utilization initiative. With a fluvial formation geology and large contrasts in permeability, this is a challenging and novel application of CO2 enhanced oil recovery (EOR). We build a predictive dynamic model of the subsurface that incorporates the multiphase and compositional data acquired during the pilot operation. The calibrated model is used for the carbon dioxide plume imaging. The study began with an initialization of the pilot sector model extracted from a calibrated full-field model. The pilot model calibration follows a two-step hierarchical workflow. First, we performed a large-scale update of the permeability distribution by integrating available bottomhole pressure and multiphase production data. In the second step, local permeability field is fine-tuned using a streamline-based method to match CO2 breakthrough times at the producers. The predictive capability of the calibrated model was verified through two blind validation tests: (1) the model showed good agreement with saturation logs acquired at two observation wells; and (2) the model reproduced the CO2 recovery as a fraction of the injected CO2.

The use of seismic onset times has shown great promise for integrating near-continuous seismic surveys for updating geologic models. In this study, we analyze the impact of seismic survey frequency on the onset time approach aiming to extend the application of onset time to infrequent seismic surveys. In addition, we quantitatively examine the nonlinearity of the onset time method and compare it to the commonly used amplitude inversion method. We carry out a sensitivity analysis of seismic survey frequency based on the complete seismic survey data (over 175

surveys) of steam injection in a heavy oil reservoir (Peace River Unit) in Canada. Our results show that an adequate onset time map can be obtained from the infrequent seismic surveys by interpolation between seismic surveys as long as there is no change in the dominant underlying physics between the successive surveys. The study also shows that nonlinearity of the onset time method can be - smaller than that of the amplitude inversion method by several orders of magnitude. Application to the Brugge benchmark case shows that the onset time method obtains comparable permeability update as the traditional seismic amplitude inversion method with faster computation and improved convergence characteristics.

We extend the streamline-based data integration approach to incorporate distributed temperature sensor (DTS) data using the concept of thermal tracer travel time. Then, a hierarchical workflow composed of evolutionary and streamline methods is employed to jointly history match the DTS and pressure data. Finally, CO2 saturation and streamline maps are used to visualize the CO2 plume movement during the sequestration process. The hierarchical workflow is applied to a carbon sequestration project in a carbonate reef reservoir within the Northern Niagaran Pinnacle Reef Trend in Michigan, USA. The monitoring data set consists of distributed temperature sensing (DTS) data acquired at the injection well and a monitoring well, flowing bottom-hole pressure data at the injection well, and time-lapse pressure measurements at several locations along the monitoring well. The history matching results indicate that the CO2 movement is mostly restricted to the intended zones of injection which is consistent with an independent warm-back analysis of the temperature data.

In addition to employing simulation models and inverse methods for CO2 plume imaging, we also initialized a data-driven technology for detecting inter-well connectivity based on production and pressure data. Our machine-learning framework is built on the statistical recurrent unit (SRU) model and interprets well-based injection/production data into inter-well connectivity without relying on a geologic model. We test it on synthetic and field-scale CO2 EOR projects utilizing the water-alternating-gas (WAG) process. The validation of the proposed data-driven inter-well connectivity assessment is performed using synthetic data from simulation models where inter-well connectivity can be easily measured using the streamline-based flux allocation. The SRU model is shown to offer excellent prediction performance on the synthetic case. Despite significant measurement noise and frequent well shut-ins imposed in the field-scale case, the SRU model offers good prediction accuracy, the overall relative error of the phase production rates at most producers ranges from 10% to 30%. It is shown that the dominant connections identified by the data-driven method and streamline method are in close agreement.

Texas A&M University, the lead organization in the project, was primarily responsible for the development of tomographic approaches for CO_2 plume mapping in conjunction with distributed pressure, temperature and seismic onset time data. Battelle, as a subcontractor, was primarily responsible for the development of analytical and empirical methods for analyzing transient injection rate and pressure data from point/line sources such as injection and monitoring wells. An additional area of emphasis for Battelle was the use of machine learning for such tasks as inferring reservoir connectivity information from injection-production data, and identifying variable importance for machine learning-based proxy models developed from full-physics simulations. The two organizations also collaborated on the application of the tomographic inversion methodology for a field data set. The contributions of Battelle as part of this project are documented in Appendix-B.

1.0 Introduction

The US Department of Energy (DOE)'s ongoing research program on "Safe and Permanent Geologic Storage of CO_2" has resulted in significant advances in our understanding of site characterization, modeling and monitoring technologies for CO_2 sequestration in deep saline aquifers and depleted oil and gas **fields.** One key area of focus within this R&D program has been the development of robust and cost-efficient monitoring technologies and protocols for tracking CO_2 plume migration in the subsurface. Under this funding opportunity announcement (DE-FOA-0001725), DOE is seeking the development of technologies for delineating CO_2 plumes through monitoring tools and techniques, uncertainty quantification in plume boundary detection and field validation of the proposed methods. To that end, this project seeks to ***develop and demonstrate a robust methodology for spatio-temporal tracking of CO_2 plumes during geologic sequestration based on the principle of joint tomographic inversion of seismic, pressure and temperature data***. Our motivation stems from the challenges with conventional approaches to CO_2 plume mapping using geophysical monitoring and modeling based workflows. The former is expensive, and best suited for relatively low resolution and mostly qualitative estimates of CO_2 saturation distributions, whereas the latter suffers from non-uniqueness and data sufficiency issues.

Our project addresses these shortcomings by means of a novel data integration workflow that hinges upon joint inversion of data from multiple monitoring sources. Our approach utilizes interpreted seismic onset time data at periodic intervals), as well as more frequently monitored downhole pressure and temperature measurements, to produce a quantitative assessment of the CO_2-saturation distribution at a higher resolution and with lower uncertainty than with seismic data alone. Our workflow also employs novel forward and inverse modelling strategies based on the concept of pressure and temperature "arrival times" which result in significant computational-efficiency gains. The methodology will be designed to handle both CO_2-oil-brine systems (i.e., depleted oil fields) and CO_2-brine systems (i.e., saline aquifers), and will be validated using field data from ongoing CO_2 injection projects. **Figure 1** provides an overview of how the proposed streamlined workflow from our research compares to the conventional workflow based on geophysical and pressure monitoring.

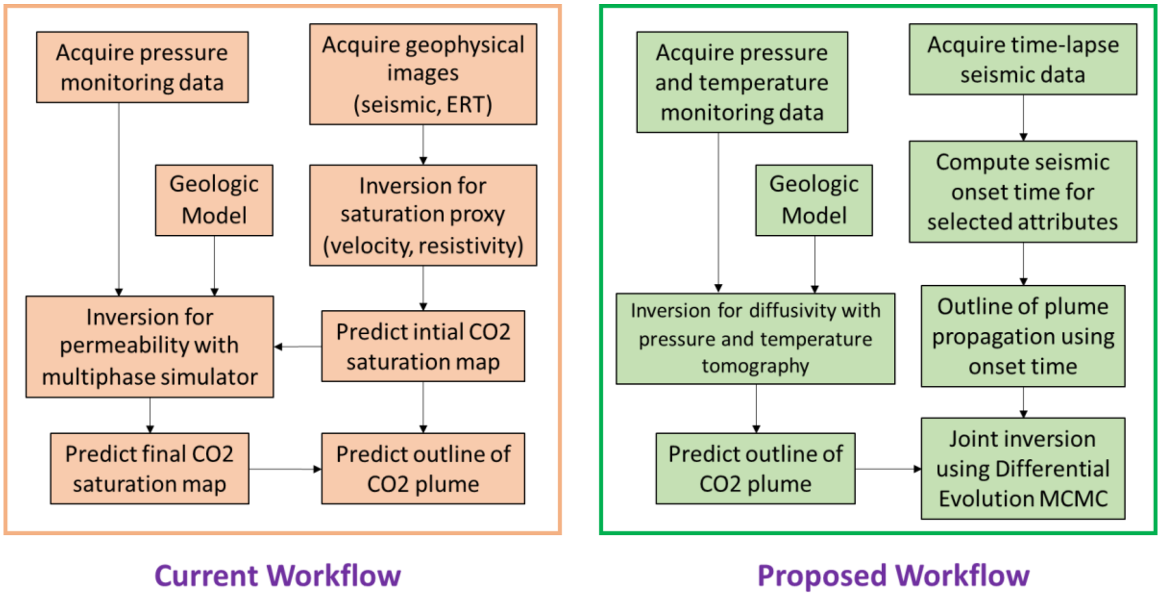


Figure 1. Overview of current and proposed workflows for CO_2 plume monitoring.

1.1 Background

Class VI Underground Injection Control (UIC) regulations require both direct and indirect site monitoring to ensure the integrity of CO_2 storage and protection of drinking water sources. As per Environmental Protection Agency (EPA) requirements under the UIC Program for CO_2 Geologic Sequestration Wells, operators must perform both direct pressure monitoring at injection and monitoring wells, and indirect monitoring and modeling of the CO_2 plume. Indirect monitoring methods targeted at tracking CO_2 plume movement include "*seismic, electrical, gravity, or electromagnetic surveys and/or down-hole CO_2 detection tools*" (EPA 40 CFR Part 146.90). Information from direct pressure monitoring is used to calibrate reservoir models, from which the spatial extent of the plume can be inferred. An overview of these two broad classes of approaches, i.e., (a) geophysical imaging based CO_2 plume delineation, and (b) reservoir modeling based CO_2 plume delineation, is described next, along with recent work on combining pressure and temperature monitoring with tomographic inversion, which has the potential to overcome the challenges of the first approach while leveraging the second approach.

(a) Geophysical imaging based CO_2 plume delineation

One successful and well established technique for monitoring and verification in CO_2 storage reservoirs is the use of seismic imaging. 3D surveys repeated in time (i.e., “4-D”) cover the largest spatial scale and its robustness and imaging strength has been demonstrated at field sites. Advanced seismic processing techniques are required to image the migration of the CO_2 within the storage complex. Time-lapse (4-D) seismic surveys have been successfully applied to image CO_2 plumes in several CCS field projects including Sleipner (Chadwick et al., 2010), Ketzin (Ivandic et al., 2015), and Cranfield (Zhang et al., 2013).

As CO_2 replaces saline water in saturated sandstone reservoirs, a P-wave velocity reduction may occur in the inter-well region. This velocity change can potentially be used to monitor CO_2 in sandstone aquifers using seismic tomography based on inversion of travel times or waveforms in order to infer CO_2-induced velocity changes in the observation plane between the wells. Cross-well seismic and VSP (Vertical Seismic Profiling) surveys have been carried out in the Frio project (Daley et al., 2008), Michigan Basin project (Gerst et al., 2013), and Ketzin project (Bergman et al., 2014), among others.

Geo-electrical methods are suited for monitoring CO_2 injected into deep saline aquifers because the electrical bulk resistivity of the medium is highly sensitive to compositional changes of the pore-filling fluids (Ramirez et al., 2003). As an example, monitoring of CO_2 plume at the Ketzin site in Germany was carried out using crosshole and surface-downhole ERT with a permanently installed vertical electrical resistivity array (Schmidt-Hattenberger et al., 2012). In addition, the inverted resistivity data were used to derive CO_2 saturation estimates, using petrophysical data from laboratory experiments. In addition, fiber optic Distributed Acoustic Sensing (DAS) is a relatively recent technology that uses an optical fiber cable as a sensor for acoustic signals. A series of field tests have been recently conducted, including: (1) CO_2 storage monitoring pilots within an oil field in Citronelle, Alabama, (2) Otway sequestration pilot project near Warrnambool, Victoria, Australia, and (3) Ketzin CO_2 pilot storage site.

Geophysical monitoring methods provide an indirect means of mapping the CO_2 plume based on a surrogate measure (e.g., sonic velocity, gravity, resistivity) and its projected response to fluid saturation changes based on some theoretical rock physics model which can have considerable uncertainty. In addition, the methods have varying degrees of spatial resolution, as well as cost of implementation. Additional constraints on such methods include technical and economic challenges if: (a) the CO_2 footprint is spatially extensive, (b) CO_2 is preferentially retained in thin high-permeability zones, or (c) CO_2 is moving in zones with insufficient sonic velocity contrast. Under such conditions there is no reliable technique currently available for imaging CO_2 plumes.

(b) Modeling-based CO_2 plume delineation

Modeling of CO_2-brine flow for geologic sequestration is commonly performed using research simulators, e.g., TOUGH2 (Pruess et al., 2011) and STOMP- CO_2 (White et al., 2012) or commercial reservoir simulators from the oil industry, e.g., CMG-GEM and ECLIPSE. The outcomes of a "forward model", given a geologic description of the subsurface, are fluid pressure and phase saturation at each grid block in the model domain. In the "inverse modeling" step, the goal is to adjust the geologic description in order to match observed values of time-dependent dynamic variables such as pressure, fluid saturation, time-lapse seismic and surface deformation data. Often, fluid saturation data are directly not measured. Therefore, calibration of the model to pressure and seismic data (if available) becomes an indirect means of inferring the extent of the CO_2 plume migration based on the simulator output values of fluid saturations.

One of the key challenges with a modeling-based approach to plume tracking is that the pressure response is influenced by diffusivity (i.e., ratio of permeability to compressibility), where the saturation prediction is influenced by advective characteristics (e.g., intrinsic and relative permeability). There is considerable uncertainty in defining field-scale relative permeability curves, which are typically measured only at the lab scale. Mishra et al. (2014) showed that non-unique combinations of intrinsic and relative permeability coefficients can produce similar pressure match, but different estimates of plume migration (as depicted in Figure 2). In practice, only a few inverse modeling studies can be expected to be performed during the course of a project because of data and computation needs. Manual history matching can be time consuming, cumbersome and difficult to assess. Traditional inverse methods have relied on numerical perturbation or adjoined-based sensitivity calculations which can be computationally intensive and difficult to implement because of the complexity of the algorithms and the requirements of access to the source code. Thus, only a limited number of snapshots of the spatial extent of the CO_2 plume can be realistically obtained from the periodic inversion of pressure data. The confidence with modeling-based plume tracking is also likely to be impacted by issues of non-uniqueness (i.e., same pressure match leads to different plume images) and resolution (i.e., good pressure match does not automatically imply a good saturation match) as noted above. Therefore, some of the potential areas of improvement with the modeling-based approach can be identified as: (a) better resolution and lower uncertainty in plume delineation, (b) increased frequency, i.e., (near) real-time, in conversion of pressure data to spatial plume extent, and (c) computational efficiency for routine field-scale applications.

(c) Pressure and temperature tomography based plume delineation

Pressure tomography refers to the use of arrival time of a pressure signal from the injection to the observation well to infer characteristics of the subsurface (e.g., hydraulic diffusivity. Over the last two decades, using concepts similar to seismic inversion, hydraulic (pressure) tomography has developed as an attractive alternative for characterizing spatially distributed groundwater aquifer properties (e.g., Butler et al., 1999; Brauchler et al., 2007; Vasco and Datta-Gupta, 2016). Here,). One efficient approach for solving the corresponding inverse problem is based on the approximation of the transient groundwater flow equation by an Eikonal equation (e.g., Vasco et al., 2000; Datta-Gupta et al., 2001; Brauchler et al., 2011). This inversion produces tomograms of hydraulic diffusivity (i.e., ratio of transmissivity to storativity).

Recently, Hu et al. (2015) have suggested that this pressure tomography approach can be used to map the spatial extent of the CO_2 plume. The premise is that CO_2 injection creates a transient heterogeneity in the hydraulic properties of the system, which can then be translated into maps of aquifer diffusivity (mirroring the CO_2 plume migration) via tomographic inversion. However, they point out that quantitative estimates of the spatial distribution of CO_2 saturation are only possible by calibration with an appropriate multi-phase simulator. To make the approach computationally feasible, Hu et al (2015) used a single phase emulator to approximate the multiphase flow of CO_2 and brine. Our proposed approach expands on this idea by advancing the technology in several different areas, as will be pointed out in the next section.

Distributed temperature surveys are becoming increasingly common to understand downhole flow conditions, especially in CO_2 injection projects involving complex well/reservoir domains. The correlated movement of pressure and thermal pulses in response to CO_2 injection is well known (e.g., Pruess et al., 2001). In the past, the analysis of temperature data in CO_2 projects have been mostly limited to flow diagnostics based on semi-analytical approach (e.g., LaForce et al., 2013). For oil and gas applications, several studies have looked at the joint inversion of pressure and temperature data for inferring reservoir properties, and indirectly, the saturation distribution (e.g., Duru et al., 2010; Li et al., 2011; Ribiero and Horne, 2013). However, the use of an arrival time concept of the thermal pulse has only been explored recently, and that too in the context of forward models for two-phase flow in shale gas wells (Cui et al., 2016). We will build on this experience in the analysis of distributed temperature response to develop an efficient formalism for modeling temperature ‘front’ propagation and inversion of distributed temperature measurements along with pressure data, specifically for the problem of CO_2 plume migration.

(d) Joint tomographic inversion based plume delineation and uncertainty assessments

In addition to the unresolved issues in seismic and fluid flow modeling, challenges remain in the area of integrated inversions and uncertainty assessments. Uncertainty is inherent in dynamic reservoir modeling because of several factors, the primary ones being the uncertainty in geologic models, errors in forward modeling and data noise. The uncertainty in reservoir parameters is translated into uncertainty in reservoir performance for CO_2 sequestration that will impact the economic and operational risk analysis. In the context of the Bayesian inversion, the solution to the inverse problem is the posterior probability distribution itself. Therefore, the problem of uncertainty quantification is closely tied to the correct sampling from the posterior distribution (Efendiev et al., 2008; Ma et al., 2008; Mondal et al., 2010). Such sampling is nontrivial because the posterior distribution is defined on a high dimensional space and is not known in a closed form. Furthermore, the posterior distribution can be both non-Gaussian and multimodal. This makes rigorous sampling from the posterior distribution extremely computationally demanding.

Another challenge is the diverse forms of fluid flow and seismic data that can be potentially conflicting, particularly because of the interpretative nature of the seismic data. We plan to explore the use of multi-objective algorithms for probabilistic integration of diverse data types. Use of fast flow simulation for rapid likelihood computation and faster convergence will be critical to the practical feasibility of our approach.

1.2 Objectives of the Project

The goal of this project is to develop and demonstrate a rapid and cost-effective methodology for spatio-temporal tracking of CO_2 plumes during geologic sequestration using joint tomographic inversion of seismic data and distributed pressure and temperature measurements. Key elements of our methodology are: (a) a computationally efficient approach to pressure and temperature propagation calculations using the Fast Marching Method (FMM), (b) analysis of time lapse seismic data using a novel ‘seismic onset time’ approach to detect fluid front propagation , (c) data assimilation and uncertainty assessment via joint Bayesian inversion of pressure, temperature and time lapse seismic data, and (d) field validation of the methodology using data collected from ongoing field projects involving CO_2 utilization for enhanced oil recovery and frequent time lapse seismic monitoring. The project team is led by Texas A&M University and includes Battelle as a subcontractor with support from Shell, Anadarko, Chevron and JX Nippon.

The project will develop a robust and cost-effective methodology for indirect imaging of CO_2 plume using interpreted seismic onset time data at periodic intervals, as well as more frequently monitored distributed pressure and temperature measurements, in conjunction with forward and inverse modeling strategies that are both novel and computationally efficient. An Eikonal formulation for pressure and temperature propagation during CO_2 injection will be developed and solved using the Streamlines and Fast Marching Method, making it well-suited for field-scale applications. Using the propagation time, ‘the time of flight’, as a spatial coordinate, the 3-D compositional flow equations will be reduced to 1-D equation leading to orders of magnitude faster flow simulation and CO_2 saturation calculations. A robust and efficient tomographic inversion scheme will be developed for determining the spatial distribution of two-phase hydraulic diffusivity from arrival time of pressure and temperature at observation wells. This will lead to time-lapse delineation of migrating CO_2 plume using the diffusivity discontinuity concept. The seismic data integration will be carried out based on ‘onset times’ – the calendar time at which a specified seismic attribute deviates from its background value. The approach is particularly well suited when we have frequent seismic surveys as multiple surveys are reduced to a single spatial map of onset times. This will speed up the inversion significantly. For data integration from multiple sources and uncertainty quantification in detecting plume boundaries, we will adopt a Bayesian framework using multiobjective algorithms to define a Pareto front representing the trade-off between multiple data sets during history matching. The methodology will be designed to handle both CO_2-oil-brine systems (i.e., depleted oil fields) and CO_2-brine systems (i.e., saline aquifers), and will be validated using field data from ongoing CO_2 injection projects. The outcome of the proposal will advance CO_2 plume mapping protocols using novel forward and inverse modeling techniques to: (a) reduce cost and uncertainty, (b) satisfy regulatory requirements, (c) provide continuous monitoring and long-term durability, and (d) cover a large area with improved accuracy.

The fundamental impact of this research would be to provide a practical and cost-effective methodology for CO2 plume delineation using routine pressure and temperature measurements together with seismic data. This will facilitate (near) real-time monitoring of CO2 plumes in field projects needed to meet current regulatory requirements which stipulate both direct pressure monitoring at injection and monitoring wells, and indirect (geophysical) monitoring and modeling of the CO2 plume. The methodology proposed here considers streamlining the fulfillment of these two requirements by using rapid tomographic inversion for CO2 plume monitoring.

1.3 Purpose of this Document

The purpose of this document is to provide an overview of the research completed under this project. A detailed report for each of the technical tasks has been prepared and uploaded to the DOE’s reporting website and are available to the reader who is interested in more details about the methods and results of the various tasks. The remainder of this final report provides short descriptions of each of the major technical task efforts and accomplishments.

2.0 Summary of Technical Tasks

2.1 CO2 Plume Tracking Using Pressure and Production Data

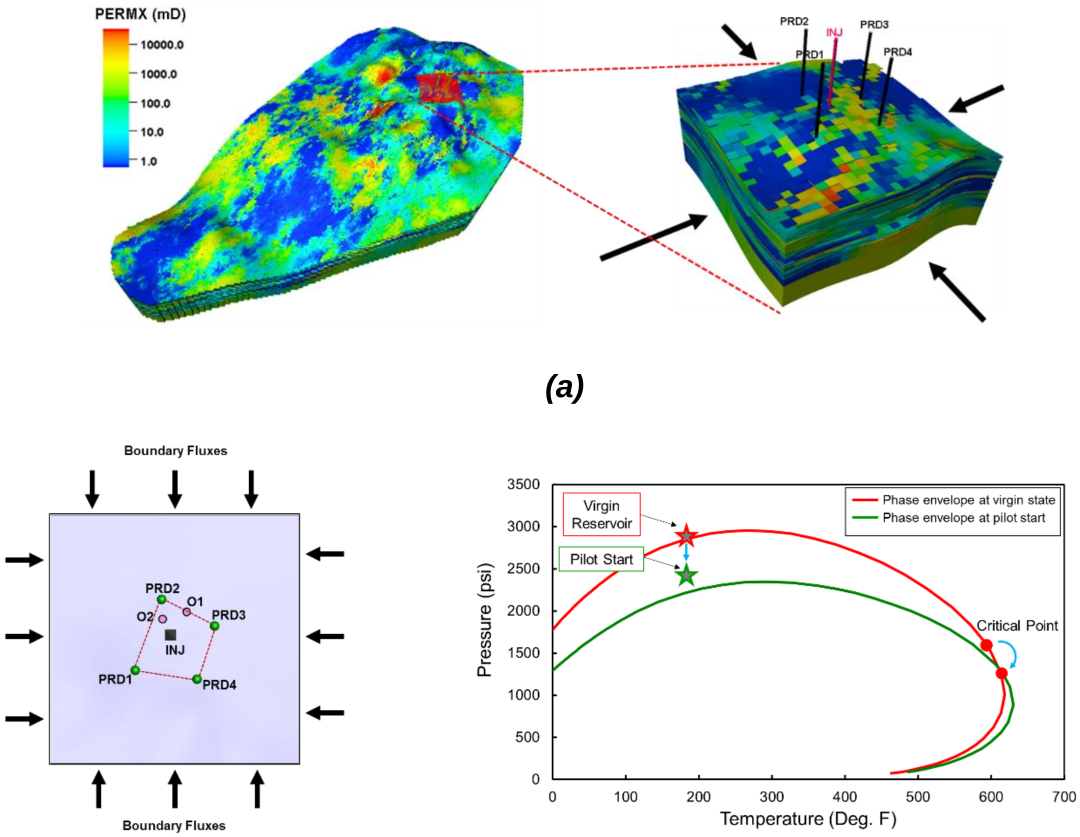
2.1.1 Introduction

The CO_2 EOR technology has been applied in different forms, based on specific nature and conditions of the oil-bearing formations for optimal oil recovery. These include continuous CO_2 injection, Water-Alternating-Gas (WAG) Injection, gravity drainage, huff-and-puff, and so on. Regardless of the process employed for CO_2 EOR, proper reservoir management plays a key role in maximizing the value of the asset using this technology (Wallace and Kuuskraa, 2014). This includes reservoir characterization to understand the subsurface flow mechanisms, followed by generation of optimal strategies for improved conformance. Measures to combat flow conformance problems become particularly vital for gas injection processes like CO_2 WAG. This is because of both viscous fingering and gravity override effects, due to differences in multiphase fluid properties including viscosity and density (Orr, 2007).

This study showcases an example of CO2 plume tracking using pressure and production data. First, we provide a background and description of the field and we also discuss the initialization of the pilot model. Second, we report the details of the pilot model calibration, starting from the model parameterization and the optimization strategy adopted to reduce data misfits in the global step of the hierarchical workflow. Next, we describe the local model update algorithm which utilizes streamline-derived arrival time sensitivities to integrate CO_2 breakthrough times into the pilot model. A discussion of the results and validation of the calibrated model using auxiliary observed data is presented next.

2.1.2 Field Description and Model Calibration

A pilot CO_2 WAG operation was conducted in a mature oil field of fluvial sandstone geology. In our previous work, multi-year field-wide production data and high-resolution downhole pressure measurements were successfully integrated into the full-field model to calibrate the reservoir energy and large-scale spatial variations of hydraulic conductivity (Olalotiti-Lawal et al., 2019). High-resolution permeability field calibration in the second stage of the hierarchical history matching workflow aided an improved reproduction of the field-wide multiphase production data by the model. Details of the field description, production history and full-field model calibration are provided in (Olalotiti-Lawal et al., 2019).



(a)

(b)

(c)

Figure 2. (a) Pilot model (right) initialization from the full-field high-resolution geologic model (left). (b) Top view of pilot model (c) Fluid phase envelopes comparison at virgin reservoir condition and current state.

For this work, the full-field model was converted to a compositional model. A sector model subsuming the pilot location was generated from the calibrated full-field model as shown in Figure 2 (a). The pilot CO_2 WAG operation was conducted with an inverted 5-spot pattern flooding. Two observation wells O1 and O2, shown in Figure 2 (b), were used to acquire reservoir saturation logs to monitor CO_2 plume movement during the pilot period. To ensure physical consistency, phase and component fluxes at the pilot boundaries throughout the pilot test period were mapped from the full-field onto the pilot model, as shown in Figure 2 (a) and (b). Pressure field, phase saturations and component mole fractions were initialized directly from the full-field profiles obtained from the compositional simulation of the multi-year history of the reservoir. As expected, due to variation in fluid compositions through the long production history of the field, moderate changes in fluid phase behavior can be noticed between reservoir fluids at virgin conditions and at the current state. This is shown in comparison of the phase envelopes (pressure-temperature diagrams) in Figure 2 (c).

The CO_2 WAG pilot entails 19 days of CO_2 injection followed by 1 year of water injection. The time series data sets acquired during this operation include multiphase production data, CO_2 mole fraction at the producers and wellhead pressure data. For all producers and injector, the bottomhole pressure data was estimated from the wellhead pressure data using a multiphase vertical lift model.

We follow a two-step hierarchical approach for the calibration of the pilot model (Yin et al., 2011) as described in the general workflow schematic in Figure 3. In the first step, we focus on a global update of spatial distribution of the model permeability field and the relative permeability functions. We apply the Grid Connectivity Transform (GCT) which enables spatial updates of the permeability field using a small set of basis functions (Bhark et al., 2011; Olalotiti-Lawal and Datta-Gupta, 2019). In the second step, we fine-tune the pilot model permeability field by matching CO_2 arrival times at the producers using the streamline-based Generalized Travel Time Inversion (GTTI) algorithm (Cheng et al., 2005; Cheng et al., 2004; He et al., 2002).

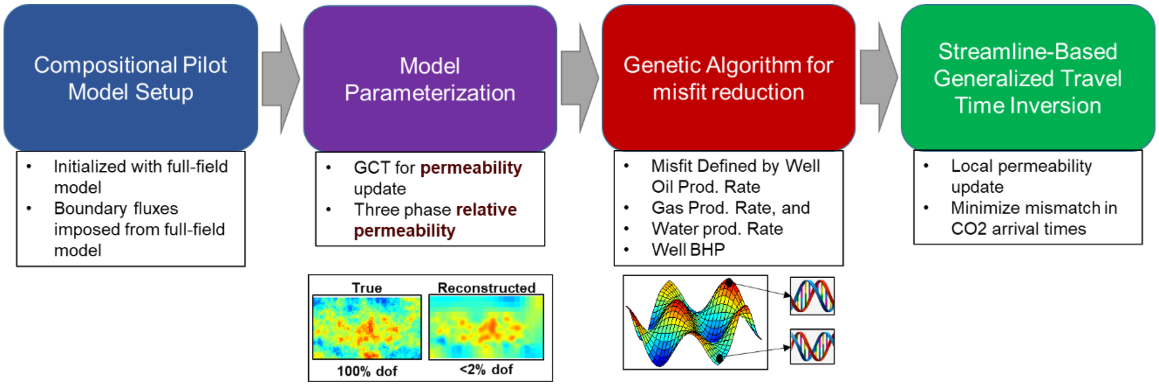
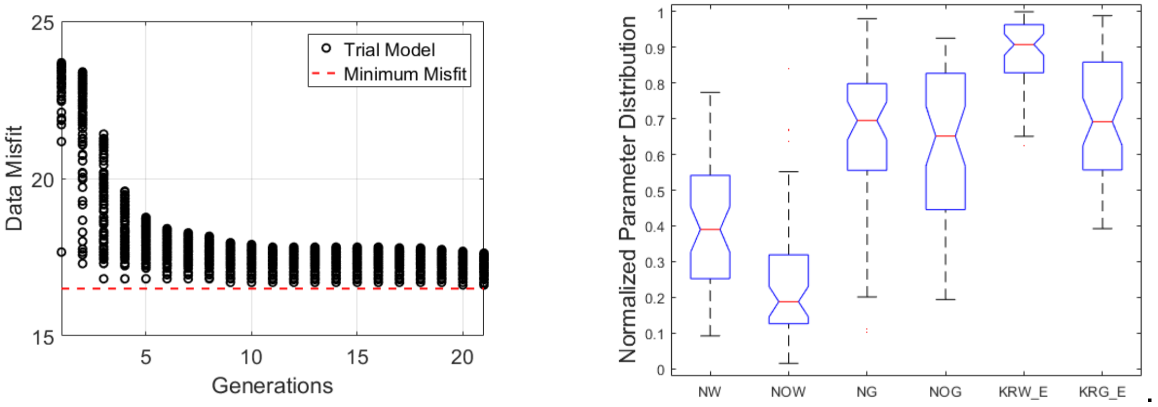


Figure 3. Hierarchical pilot model calibration workflow

2.1.3 Model Calibration Results and Validation

A plot of the data misfit reduction obtained from GA in the global model calibration step is shown in Figure 4 (a). Although, the 'best' model is selected as the one with the least data misfit at the end of the 20 generations, we also provide a boxplot of the normalized distribution of the parameters in the 80 members of the population in the last generation in Figure 4 (b).



(a) ***(b)***

Figure 4. Global update results showing (a) data misfit reduction with generation and (b) normalized distribution of updated parameters

Production matches obtained from the pilot model calibration are summarized in Figure 5. The initial model response, the model response after the global update and after the local update are compared with the observed data. The results show improved matches for all 4 wells in terms of gas production rate.

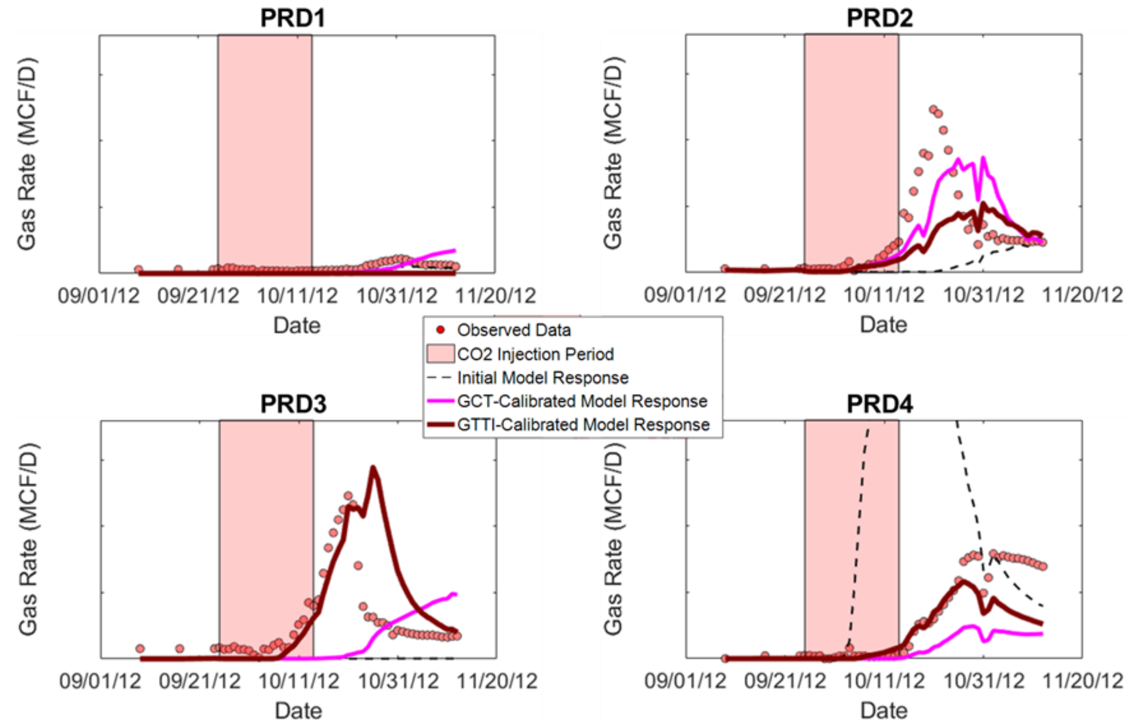


Figure 5. Comparison of Gas production Rates (GPR) at each stage of the model calibration (plot-scale suppressed per operator request).

It is expected that a significant portion of injected CO_2 is trapped in the subsurface and the proportion recovered provides an indication of the quality of sweep in the subsurface (Melzer, 2012). A field-wide comparison of CO_2 recovery between the initial and updated models is shown in Figure 6. An early CO_2 breakthrough is obtained from the initial model response and hence, CO_2 recovery is overestimated. This behavior can be attributed to the preferential flow path of CO_2 toward PRD4, as shown in the production responses in Figure 5, resulting in early breakthrough of CO_2 and overall poor CO_2 sweep in the subsurface. The CO_2 recovery was brought within the range of the observed data in the field in the calibrated pilot model. The updated model response also showed good agreement with the observed data in terms of CO_2 breakthrough time.

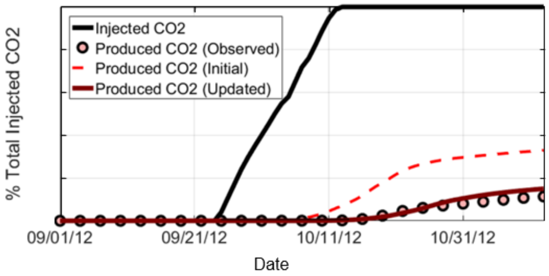


Figure 6. CO$_2$ recovery fraction comparison between observed data and initial and calibrated model responses.

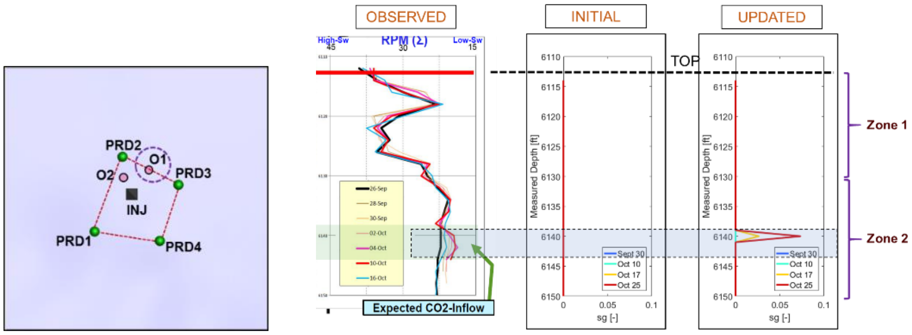


Figure 7. Reservoir saturation logs (sigma responses) comparison between observed data, initial model response and calibrated model response at observations well O1

Next, we compared the model response with gas saturation logs at an observation well O1. The locations of the observation well O1 relative to the other wells in the pilot pattern are shown in Figure 7. The observed data is compared with the gas saturation profiles at the wells. The wireline measurements serve as proxy for reservoir saturation logs (sigma responses) so that time variations in the measurements at specific zones down the observation well indicate gas invasion through that zone. CO$_2$ is expected to flow into Zone 2 of the reservoir at the location of O1 according to the observed data. While this could not be reproduced by the initial model, the profile is correctly captured in the updated model.

2.1.4 CO2 Plume Propagation

In Figure 8**,** visual comparisons of the CO$_2$ plume between the initial and updated models at the end of CO$_2$ injection are provided. In all the plots presented, the plume shape was obtained at a threshold value of 0.3 CO$_2$ model fraction. As shown in Figure 8, CO$_2$ never flowed towards PRD3 due to the poor hydraulic connectivity with INJ in the initial model. This was corrected in the updated model and the simulation response showed CO$_2$ breakthrough at PRD3 as in the observed data. Overall, while CO_2 sweep appears to be improved in the updated model compared to the initial one, both models show preferential CO_2 flow within Zones 1, 2 and 4 and practically no flow in Zone 3.

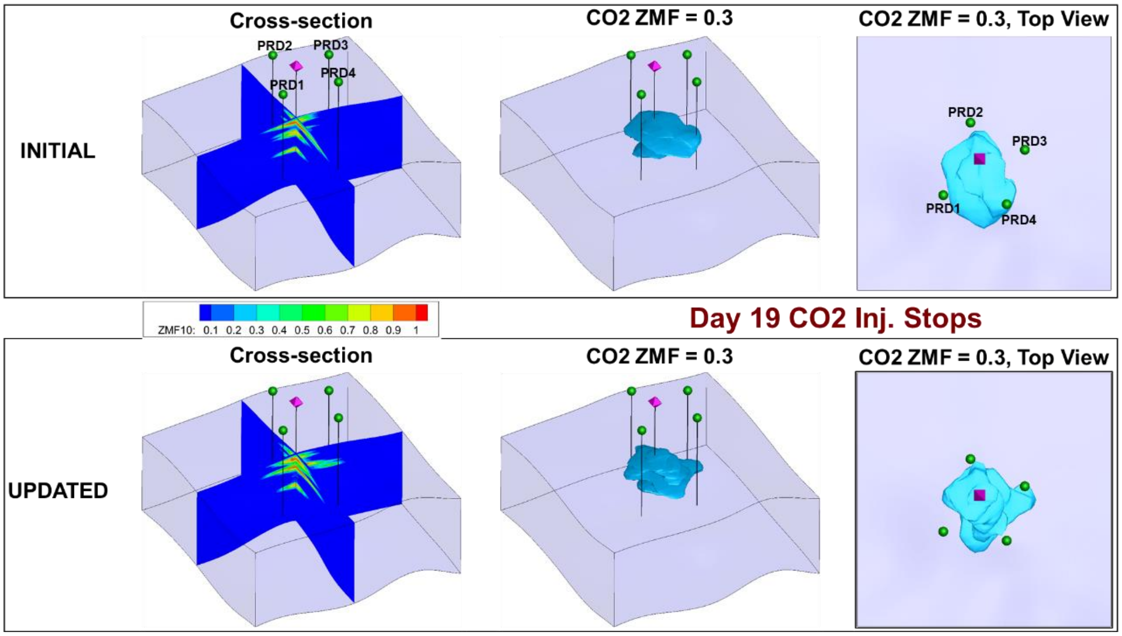


Figure 8. Visual comparison of CO2 plume migration between initial and calibrated models.

2.2 Seismic Time-Lapse 'Onset' Times for Imaging Front Propagation

2.2.1 Introduction

With the development of seismic monitoring technology, the use of time-lapse (4D) seismic data for characterization of subsurface flow has become increasingly popular. A number of successful field applications of reservoir management based on 4D seismic data have demonstrated its power and efficacy (Behrens et al., 2002; Fahimuddin et al., 2010; Landrø et al., 2001; Landrø et al., 1999; Watanabe et al., 2017; Hetz et al., 2017). The integration of 4D seismic data into high-resolution geologic models usually involves least-squares-based minimization to match the simulated seismic response and the observed seismic signals (Dadashpour, Mohsen et al., 2010; Dadashpour et al., 2009; Falcone et al., 2004; Gosselin et al., 2003; Gosselin et al., 2001; Rey et al., 2009; Rey et al., 2012; Rwechungura, R.W. et al., 2012; Vasco, 2004; Watanabe et al., 2017; Tang et al., 2018). The objective function for the minimization has mostly focused on matching the seismic amplitude, travel-time, waveform, etc. Such inverse problems are usually highly nonlinear and ill-posed and can easily cause the solution to converge to a local minimum, leading to an inadequate history match (Cheng et al., 2005). More recently, the seismic inversion method based on the seismic onset times has shown great potential for integrating 4D seismic data into high-resolution geologic models (Hetz et al., 2017; Vasco et al., 2015; Vasco et al., 2014). Vasco et al. (2014) introduced the seismic onset time as the calendar time at which a measured quantity, such as seismic travel time or reflection amplitude, begins to deviate from its background value. Using the concept of seismic onset time, multiple sets of 4D seismic data can be converted into a single onset time map, which can be used for dynamic model updating (Hetz et al., 2017).

In the previous research, the onset time approach has shown its efficacy and robustness for integrating frequent seismic surveys where the propagation of saturation front could be captured in detail with the short time span between seismic surveys. However, due to the high cost associated with conducting seismic surveys, frequent seismic surveys are usually not commonly available. For most field situations, seismic surveys are usually conducted more infrequently, typically in months or years. To overcome the lower time resolution in infrequent seismic data, we examined the use of various interpolation methods to calculate the onset time map. We also demonstrate the advantage of onset time approach for seismic data integration by comparing the onset time inversion method with the traditional seismic amplitude inversion method using the Brugge benchmark case. Moreover, the nonlinearity associated with these two methods has been quantitatively investigated.

2.2.2 Onset Time as Seismic Observation for Infrequent Seismic Data

The onset times (Vasco et al., 2015) are defined as the calendar times at which the measured time-lapse attributes begin to deviate from their initial or background values above a pre-defined threshold value. Using the onset time approach, multiple sets of time-lapse seismic data can be converted into a single onset time map, which represents the propagation of the change (change in the fluid saturation, pressure, temperature, etc.) within the reservoir.

To illustrate the onset time approach, a 2D synthetic case with five-spot waterflooding pattern is shown below (Figure 9). Consider that 5 sets of time-lapse seismic surveys were conducted over a perid of 2200 days. Every two consecutive seismic surveys are either 400 or 600 days apart, rather sparse in time. The observed acoustic impedance maps and the methodology to calculate the onset time map for these infrequent seismic surveys are shown in Figure 10.

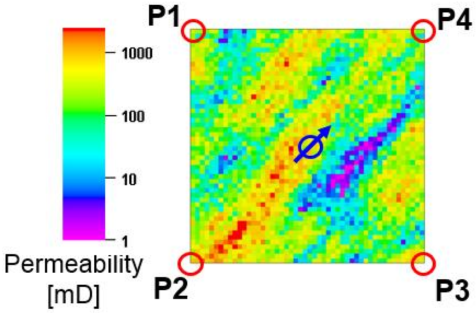


Figure 9. 2D (50x50) waterflooding case

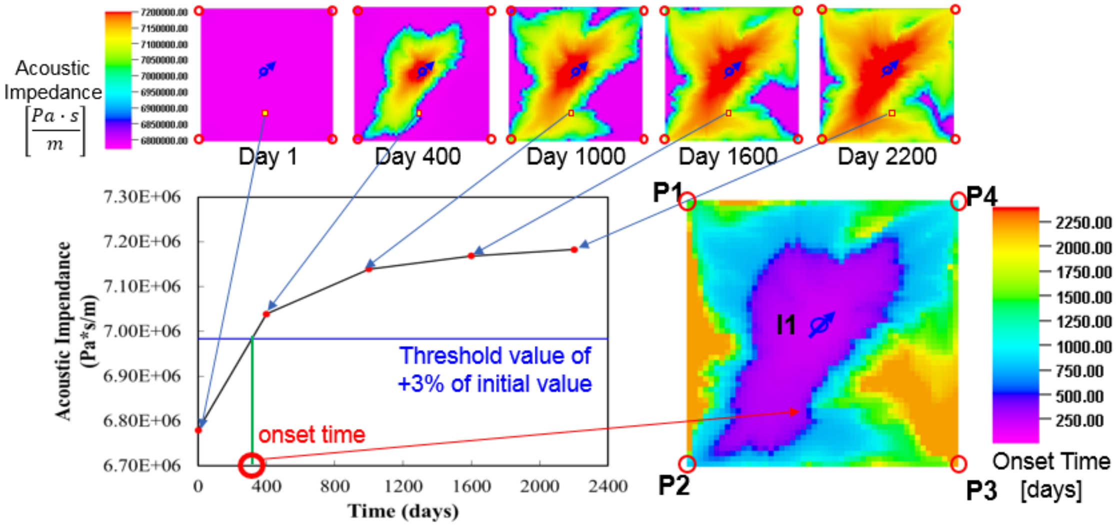


Figure 10. Onset time calculation for infrequent seismic surveys

2.2.3 Onset Time Inversion Workflow

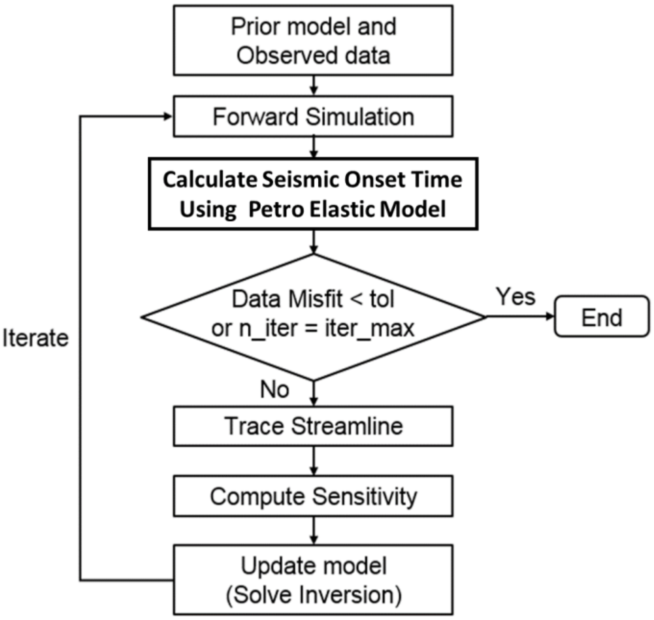


Figure 11. Onset time inversion workflow

The onset time map is a valuable representation of the propagation of change in the reservoir, which can be used for dynamic modeling and model calibration through a history matching process. The onset time inversion workflow for integrating time-lapse seismic data is shown in Figure 11. The objective of seismic inversion is to update the prior model parameter based on the observed seismic data. The seismic inversion is conducted in an iterative manner by updating the reservoir model parameters until a desired match between the simulated onset time map and the observed seismic onset time map is achieved.

2.2.4 Field Application

The Brugge field model was designed for a benchmark project to test the combined use of history matching and waterflooding optimization workflow (Peters et al., 2010). The structure of Brugge field shown in Figure 12. There are 20 producers located at the top of the dome within the oil-rich region. 10 peripheral water injectors provide pressure support in addition to the bottom aquifer. The injection rate is 10,000 bbl/day and the producers are constrained with a liquid production rate of 5,000 bbl/day.

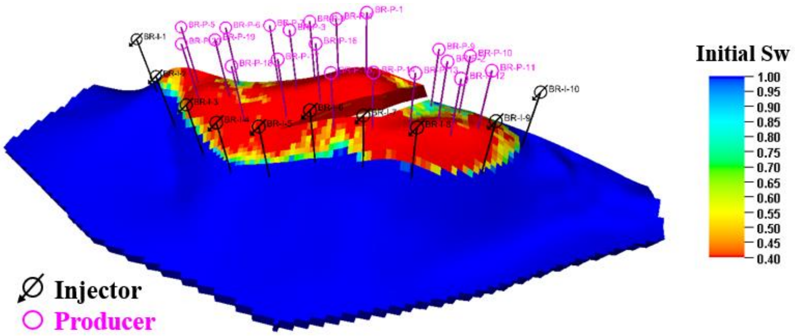


Figure 12. Initial water saturation of Brugge model with blue color showing the bottom aquifer and red color showing the oil on the top of the dome

The reference model is used to generate the observed seismic data every year since the beginning of the waterflooding process. The observed acoustic impedance maps are shown in Figure 13. Using a threshold value of an increase of 0.8% of the initial acoustic impedance, the interpolated onset time map is obtained as shown in Figure 14.

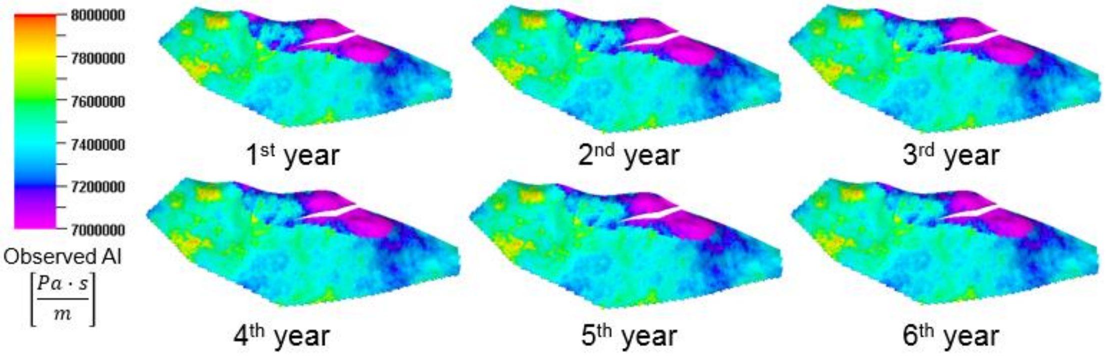


Figure 13. Observed acoustic impedance (layer 7) maps generated with reference model for 6-year waterflooding process

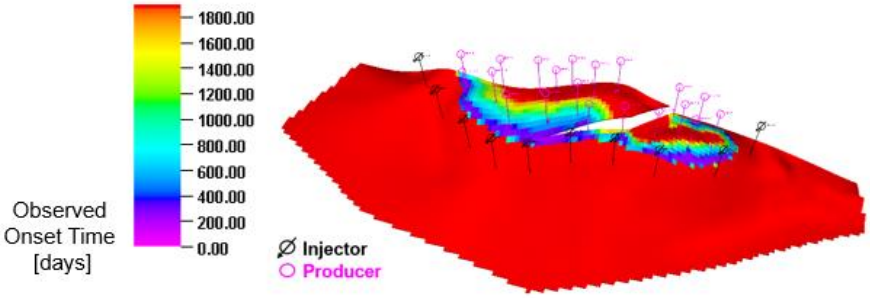


Figure 14. Observed onset time map (layer 7)

We compare the inversion performance of onset time method based on a single onset time map with the amplitude method based on 5 acoustic impedance change maps. The 5 sets of acoustic impedance change maps are the differences between consecutive acoustic impedance maps. The quality of model update, convergence performance, and CPU time are compared. Additionally, the nonlinearity associated with these two methods are quantitatively investigated.

The permeability updates for the selected layer 7 are shown in Figure 15. The updates in the model via onset time inversion and amplitude inversion are found to be similar. Both updated models give better water cut matches than the initial model. Among the 20 producers, all the wells with water breaking through show improved water cut predication than the initial model. Two wells, producer 3 and 16 are shown for illustration in Figure 16. However, the onset time approach is more efficient and converges faster than the amplitude inversion method. Figure 2Figure 17 shows the comparison of the reduction of normalized data misfit and the CPU time. The onset time inversion method converges much faster than the amplitude inversion method, obtaining 2 times more error reduction after 15 iterations. Additionally, the onset time method costs one sixth computational time of the amplitude inversion method.

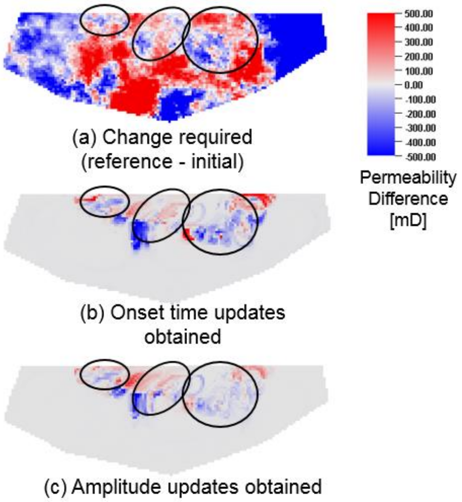


Figure 15. Comparison of permeability change (layer 7) with black circles indicating areas with similar permeability updates, (a) change required (reference - initial), (b) onset time updates obtained (onset time updated - initial), (c) amplitude updates obtained (amplitude updated - initial)

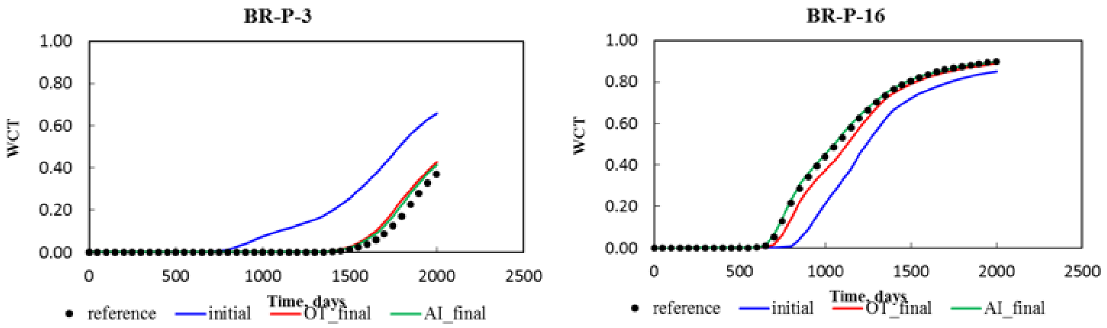
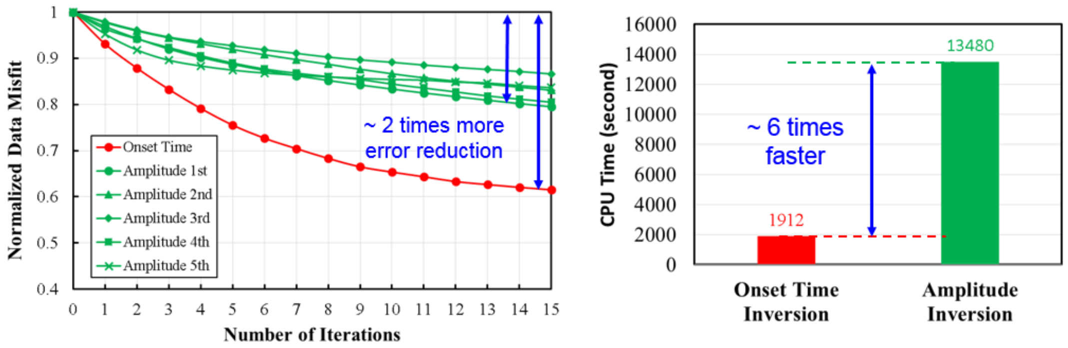


Figure 16. Comparison of water cut predication for producer 3 and 16



(a) (b)

Figure 17. Comparison of the inversion performance, (a) reduction of normalized data misfit, (b) CPU time

The superior convergence performance of onset time method results from its advantage of being more linear than the amplitude inversion method. The maximum nonlinearity values for each iteration are compared in Figure 18. Clearly, it can be seen that the nonlinearity of onset time method is not only much sparser in space but also much smaller in magnitude.

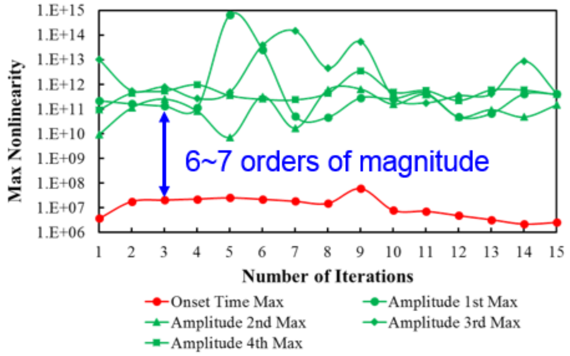


Figure 18. Comparison of maximum nonlinearity of each iteration

2.3 Data Assimilation: Using Pressure, Temperature and Production Data for Plume Imaging

2.3.1 Introduction

Injecting CO_2 into the subsurface poses certain risks, including CO_2 leakage through wells or non-sealing faults into groundwater or to the earth's surface (Ennis-King and Patterson 2002; Tsang et al. 2002; Hesse and Woods 2010). Monitoring and risk assessment of the subsurface CO_2 movement during injection and post-injection periods are necessary in these projects. Many CO_2 monitoring studies based on different types of data are found in the literature, such as time-lapse seismic (Li, G. 2003; Rey, A. et al. 2010), gravity measurements (Nooner et al. 2007), well logging (Yamaguchi et al. 2006), surface deformation via satellite imaging (Mathieson et al. 2007), distributed temperature and pressure data (Mawalkar, S. et al. 2019).

Distributed Temperature Sensoring (DTS) provides high resolution time-lapse temperature measurements along a fiber-optic line. Nunez-Lopez et al. 2014 performed a study at an onshore CO_2 injection site in the U.S. Gulf Coast that identified the arrival of CO_2 plume at monitoring wells using DTS data. Miller et al. 2016 demonstrated the ability of DTS technology for downhole temperature monitoring purposes during CO_2 storage process at Petroleum Technology Research Council's Aquistore storage site. Mawalkar et al. 2019 demonstrated a warmback analysis using DTS data for monitoring injection of CO_2 into a depleted oil reservoir. However, the inferences drawn from the warmback analysis were qualitative. This study will introduce the streamline method (Datta-Gupta and King 2007; Datta-Gupta, Xie, et al. 2011; Chen et al. 2021) to quantitatively apportion volume of CO_2 injection into various formations.

To quantitatively analyze CO2 injection and plume evolution, we first developed the streamline-based history matching method to account for the thermal process and DTS data and validated this algorithm using a 2D synthetic case. Then, the proposed method is combined with Genetic Algorithm (GA), establishing a hierarchical workflow that assimilates Distributed Temperature Sensoring (DTS) data, bottom-hole pressure (BHP) data and behind-casing pressure sensor data collected from the Department of Engergy Midwestern carbon sequestration project in Northern Niagaran Pinnacle Reef reservoir in Michigan. The history matched model is then used for plume evolution visualized by streamline.

2.3.2 Thermal Front Propagation

The transit of a neutral tracer under the influence of a velocity field can be characterized by the Time of Flight (TOF) (Datta-Gupta and King, 2007). The calculation of TOF, τ, along an arbitrary flow path or streamline ψ can be mathematically expressed as:

$$\tau = \int_{\psi} \frac{\phi(r)dr}{u(r)} \tag{1}$$

where $\phi(r)$ is the porosity, $u(r)$ is the Darcy velocity, and dr is the distance element along the streamline.

Similarly, assuming that the heat transport during a CO_2 injection process in the subsurface is dominated by advection (Lake, 1989), the propagation of a thermal tracer can be characterized by the thermal tracer TOF, whose calculation can be obtained by modifying the above equation with a thermal retardation factor (R) (Somogyvari and Bayer, 2017):

$$\tau_{tt} = \int_{\psi} \frac{\phi(r)dr}{R(r)u(r)} \tag{2}$$

where τ_{tt} is the travel time of the thermal tracer and R is the thermal retardation factor that describes the lag in the travel time of heat relative to a neutral tracer. R depends on the porosity and the heat capacities of the fluid (C_f) and the rock (C_r) ($R=C_r/(\phi C_f)$).

In this work, we incorporated the concept of thermal tracer TOF in the existing streamline-based inversion workflow, enabling the streamline method for the DTS data history matching.

2.3.3 Inverse Problem Formulation

The mathematical formulation behind the streamline-related inverse problems has been discussed in detail (Vasco and Datta-Gupta, 1999; He et al. 2002), where a penalized misfit function is defined as below:

$$|| \delta d - S\delta R || + \beta_1 || \delta R || + \beta_2 || L\delta R || \tag{3}$$

where δd is the data misfit, S is the sensitivity matrix with respect to gird properties and δR corresponds to the change in the reservoir properties, which is grid permeability in this study. The second component, called the norm constraint, penalizes deviations from the prior model, keeping the changes small to preserve the geology. The third component is the roughness constraint, which is calculated by the Laplacian matrix L and used to avoid unrealistic sharp changes applied to the geologic model. β_1 and β_1 are tuning weights applied to the penalty terms and there are guidelines in the literature for selecting them (Parker 1994). The penalized misfit function is minimized by an iterative workflow as shown in Figure 19. The inversion is conducted in an iterative manner by updating the reservoir model parameters until a desired match between the simulated DTS and the observed DTS is achieved.

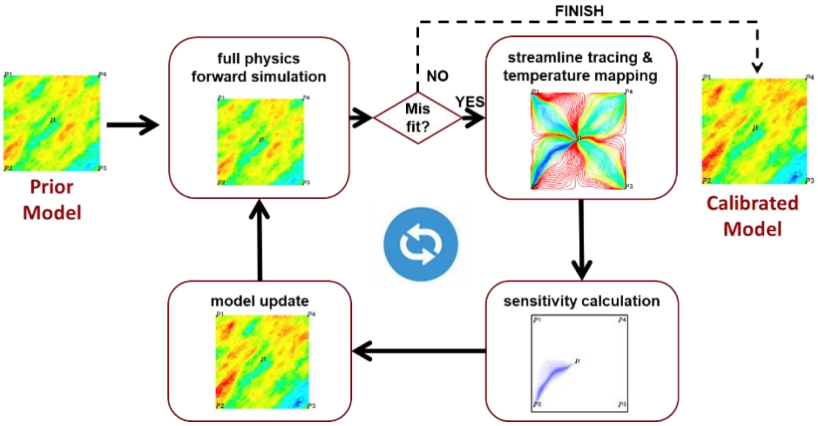


Figure 19. Diagram of streamline history matching workflow

2.3.4 Field Application

We applied our proposed DTS integration algorithm, combined with Genetic Algorithm (GA), to CO_2 injection project located at the Northern Niagaran Pinnacle Reef Trend (NNPRT) in Michigan, which is part of the DOE's Midwest Regional Carbon Sequestration Partnership (MRCSP).

The Midwest Regional Carbon Sequestration Partnership (MRCSP) was established in 2003 to assess the technical potential, economic viability and public acceptability of carbon capture, utilization and storage (Gupta et al. 2014). This CO_2 injection project is part of the MRCSP and more than 1 million metric tons of CO_2 has been injected into Niagaran pinnacle reefs. Figure 20 shows a map-view and three-dimensional view of the Chester 16 reef with one CO_2 injection well (Chester #6-16) and one monitoring well (Chester#8-16). The Chester 16 reef includes two distinct reef cores close to one another.

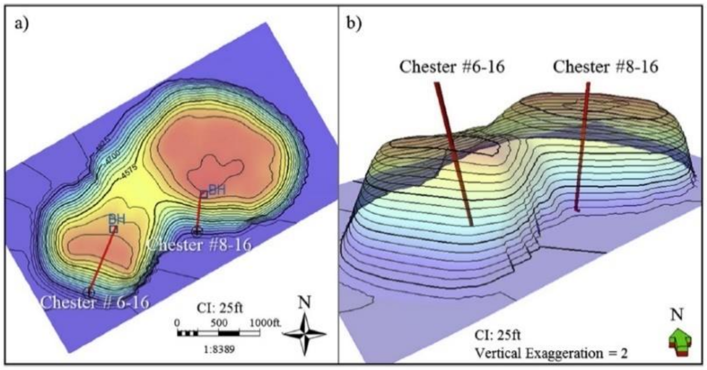
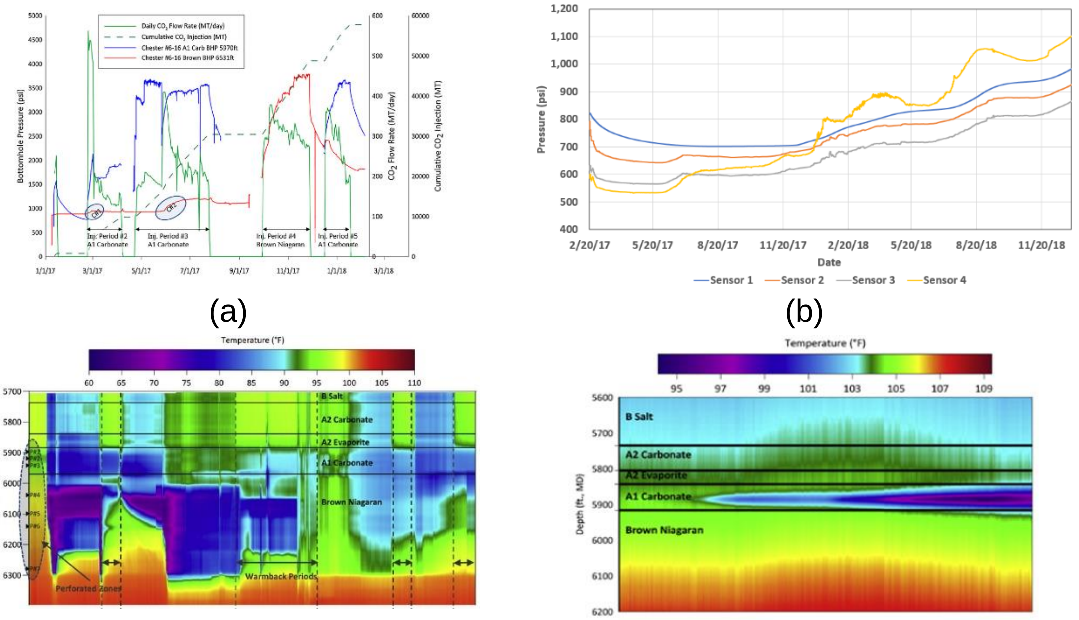


Figure 20. (a) Map-view of Chester 16 reef; (b) 3D view of the model

This study utilized a combination of pressure and temperature measurements, including bottom-hole pressure of injection well, distributed pressure measurements from four behind-casing sensors in monitoring well and DTS data of both injection well and monitoring well (Figure 21).



(a) (b)

(c) (d)

Figure 21. Observed dataset: (a) Bottom-hole pressure of injector; (b) Distributed pressure measurements from four behind-casing sensors in monitoring well; (c) DTS data at injector; (d) DTS data at monitoring well

A hierarchical history matching workflow by combining Genetic Algorithm (GA) and the proposed streamline-based inversion algorithm was applied to assimilate the above observed dataset, where distributed pressure measurements and bottom-hole pressure of injector were matched using GA by tuning several reservoir parameters selected from a sensitivity analysis and then, DTS data was integrated by the streamline-based inversion algorithm by changing grid permeability.

Pressure matching results are shown in Figure 22, where black dots denote the observed data and lines show the simulation results where the blue line is the results for the initial model, grey lines are for the best 7 realizations and the best model is highlighted by the red line. All the selected models reasonably replicate the observed well performance. The best-calibrated model will then be passed to the local calibration process to incorporate DTS data into the calibration process by changing grid permeability using streamline-based inversion approach.

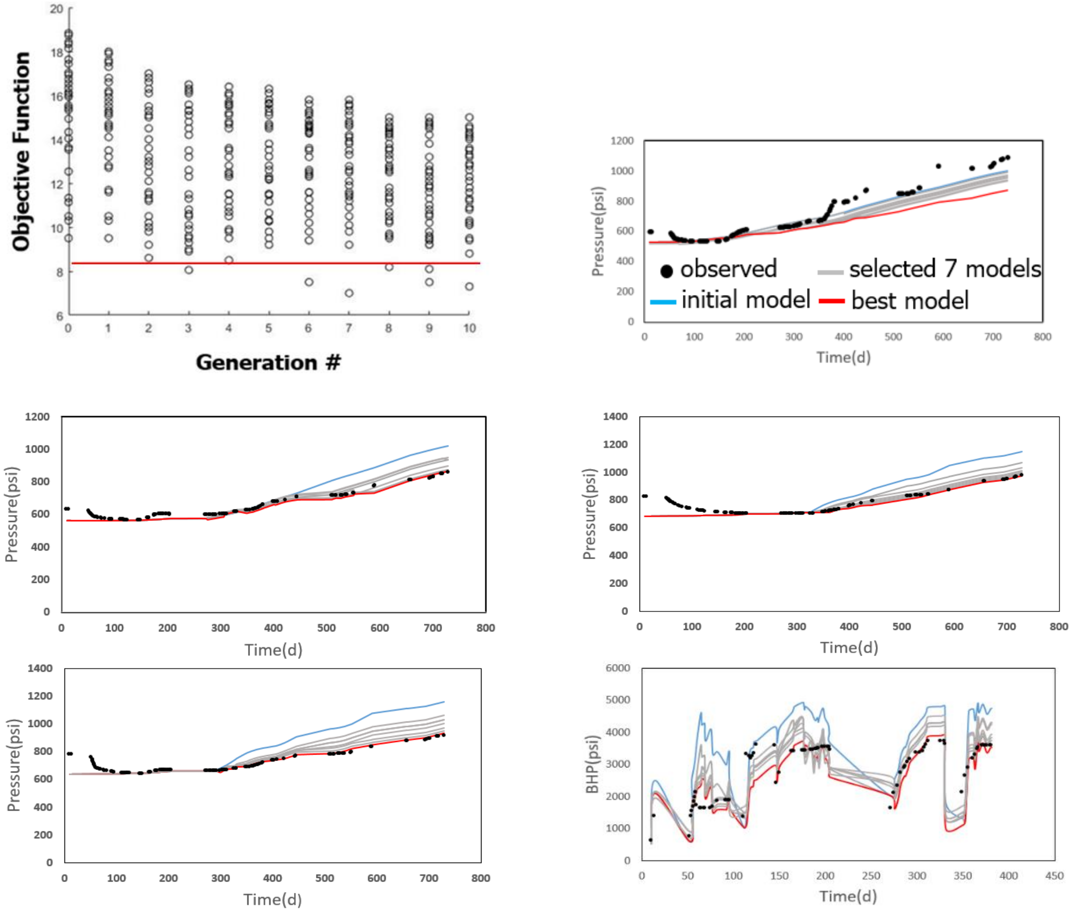
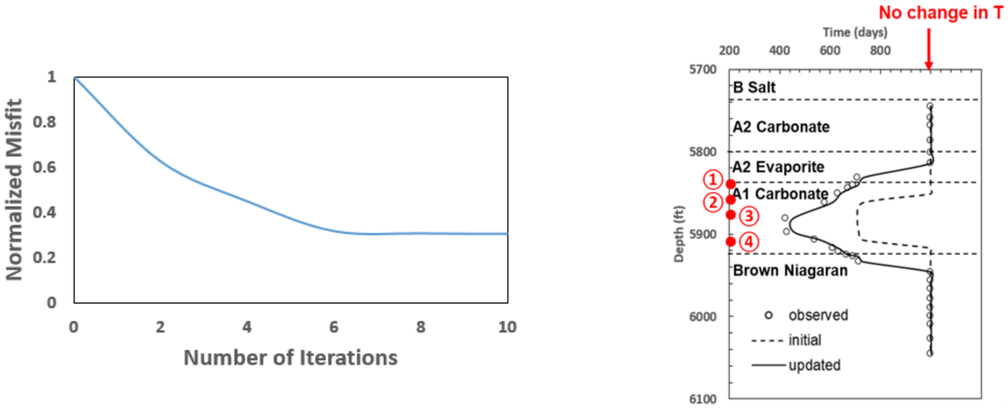


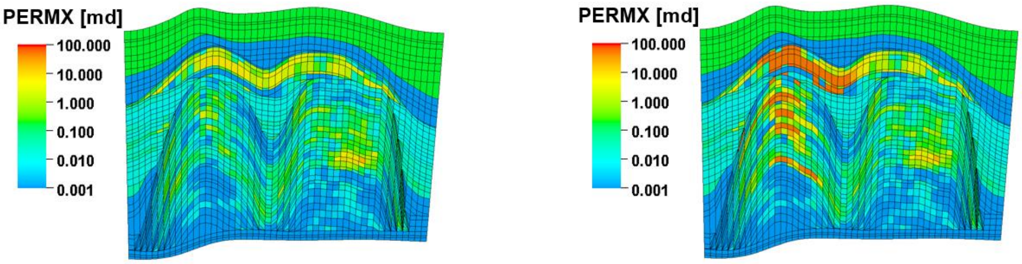
Figure 22. Pressure history matching results after GA

The DTS data inversion result is shown in Figure 23, where a significant decrease of the misfit could be observed during the inversion iterations (Figure 23 (a)), around 70% reduction after 10 iterations. The thermal tracer onset time is also calculated along the wellbore before and after model update (Figure 23 (b)). The black circle denotes the onset time calculated from the observed DTS data, dashed line is the initial simulation results, the black line is the updated simulation results and we could observe a good match between the updated simulated results and the observed data. Figure 24 shows the comparison of permeability distribution after DTS inversion.



(a) (b)

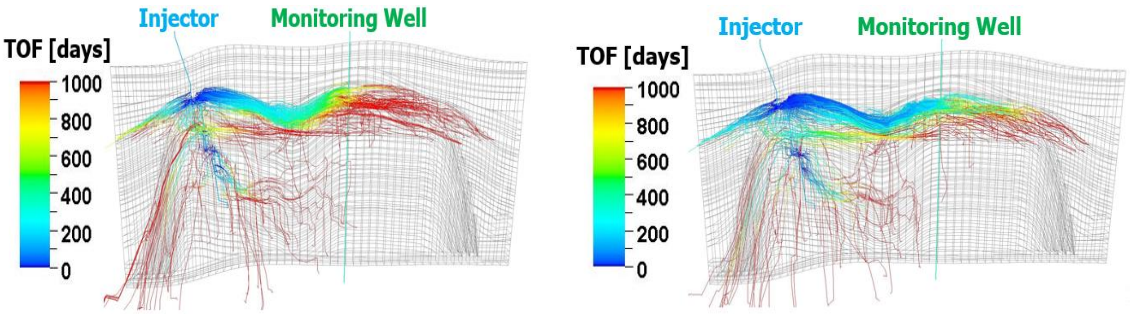
Figure 23. DTS inversion results in terms of thermal tracer onset time

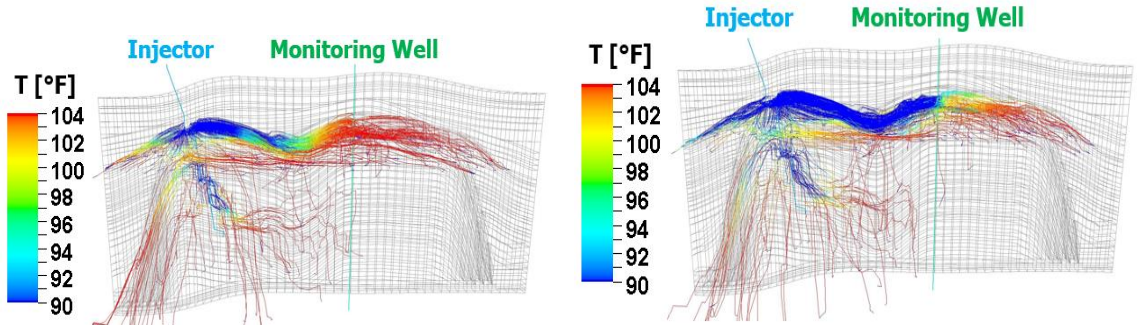


(a) (b)

Figure 24. Permeability distribution of (a) initial model; (b) updated model

Streamlines representing the CO_2 plume are traced at the end of the simulation for both the best-matched model after pressure matching and the final updated model after streamline-based DTS inversion and the results are compared in Figure 25. The streamline time of flight maps (Figure 25 (a)-(b)) and streamline temperature maps (Figure 25 (c)-(d)) show a faster CO_2 propagation and thermal tracer propagation after the history matching. Another observation is that the flow path before and after local calibration are similar, since the streamline-based inversion is designed to preserve the prior geologic model.





(a) (b)

(c) (d)

Figure 25. 3D Streamline flow path and CO_2 front visualization at the end of simulation: (a) time of flight of best-matched model from GA; (b) time of flight of updated model after DTS inversion; (c) temperature along streamline of best-matched model from GA; (d) temperature along streamline of updated model after DTS inversion

2.4 Field Application: Identifying Well Connectivities

2.4.1 Introduction

Conventional methods for assessing well connectivity include tracer tests (Zhang et al. 2016, Suarsana and Badril 2011) and numerical simulation-based techniques such as streamlines (Datta-Gupta and King 2007). Extensive literature has shown the capacity of streamlines over a variety of grid systems, including Cartesian grids (Pollock 1988), corner point grids (Cordes and Kinzelbach 1992; Jimenez et al. 2010), unstructured grids (Prevost et al. 2002; Rasmussen 2010; Zhang et al. 2011; Zuo et al. 2021), embedded discrete fracture models (Chen, Onishi, et al. 2020), dual porosity single permeability models (Chen, Yang, et al. 2020) and dual porosity dual permeability models (Chen, Yao, et al. 2020).

Major limitation of the above-mentioned works is the tremendous time cost to model general non-linearity in production trends arising from factors such as varying GOR and frequent shut-ins. A potential way to redress this shortcoming is by using a universal approximator like neural networks that can theoretically model any measurable function to an arbitrary degree of accuracy (Hornik et al. 1989). An extension of the neural network to sequential data, called the recurrent neural network (RNN), was first introduced by Hopfield (1982) and later by Rumelhart et al. (1986). From thereon, RNNs have been applied to various applications ranging from handwriting and speech recognition (Graves et al. 2008, Dutta and Sarma

2012) to regression applications such as trade forecasting (Dunis and Huang 2002) and hydrological forecasting (Coulibaly and Baldwin 2005). In the past, RNNs have been applied successfully to oil and gas problems such as formation top detection (Sen, Ong, et al. 2020), well-control optimization (Kim and Durlofsky 2021) and production prediction (Tian and Horne 2017 and 2019, Bao et al. 2020). However, connectivity detection based on RNN models is not yet demonstrated.

The statistical recurrent unit (SRU) was introduced by Oliva et al. (2017) as an un-gated alternative to more complex RNN architectures such as LSTMs (Hochreiter and Schmidhuber 1997) and GRUs (Chung et al. 2015). The SRU was shown to capture long term dependencies at least as much as the latter models, simply by using moving averages of temporal information at different scales. Therefore, the SRU architecture is simpler and more interpretable compared to LSTMs and GRUs, enabling easier training and incorporation of additional constraints such as well locations.

Variable importance is an assessment of the contribution of each input on the output (Breiman 2001). It can be a promising method of inferring well connectivity from a data-driven model. That is, the variable importance computed from an explainable data-driven model, should ideally correspond to the actual influence that an input signal (such as those from an injector) has on the output (production at producer). In this way, an explainable and predictive data-driven model may be used to derive insights on the actual reservoir connectivity in a quick and efficient way, without the need to run computationally costly reservoir simulations.

With the above in mind, we propose a SRU based framework for inferring inter-well connectivities. The proposed method contains two key components: SRU model specifically framed to the CO_2 WAG problem and the inter-well connectivity assessment based on the SRU model and variable importance calculation.

2.4.2 Connectivity Inference Workflow

The general workflow for connectivity inference is shown in Figure 26. The first step (Figure 26 (a)) is to fit a machine learning model to the data. In our application, we train an SRU model for each producer that predicts its gas production rate given the producing pressure and injector-wise injection rates. The next step (Figure 26 (b)) is to ensure the reliability of the proxy model by testing its performance on a dataset that has not been seen during training (test set). A model with good predictive power on the test set is an indication that the connections that it learned during training are representative of the true connections that exist in the field. Once we have a reliable model, we can quantify the influence of any injector on the producer by computing the permutation variable importance (Breiman 2001) which quantifies the contribution of injector's rates in the accuracy of prediction of producer’s gas production. In this way, a variable importance map can be plotted showing the major connections in the field (Figure 26 (c)).

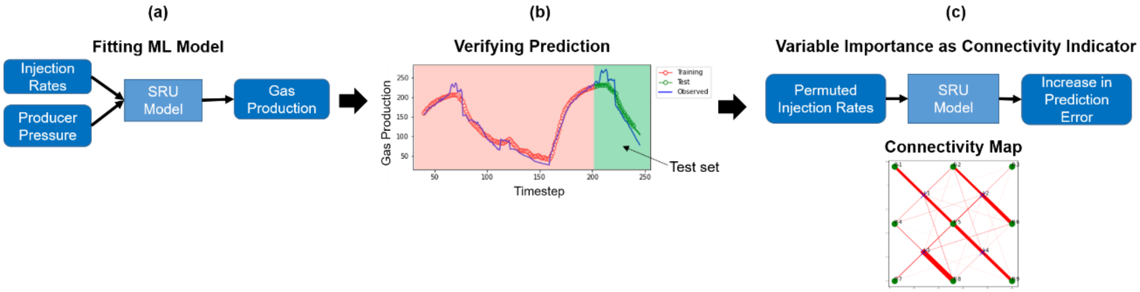


Figure 26. Connectivity inference workflow using machine learning

2.4.3 Penalty Function Accounting for Well Location Information in Deep Learning

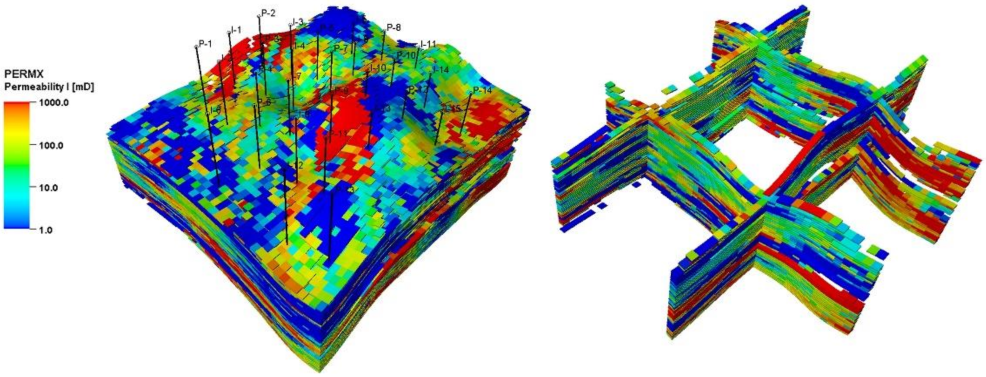
We introduce a penalty function for training the SRU in such a way that long injector-producer connections whose distance are calculated from well location information are automatically penalized during training. This way, the importance of each injector on the predicted producer is decided by the data and training process. This is realized by minimizing a loss function, given in Eq. 4, which has an extra regularization term along with the usual MSE loss.

$$Loss = \frac{1}{T}\sum_{t=1}^{T}\left((y_{obs}(t) - y_{pred}(t)\right)^2 + c\sum_{i=1}^{N_i}\left\{\sum_{j=1}^{N_{w_i}}|w_{ij}|\right\}d_i \tag{4}$$

The absolute value of these weights $(w_{i1}, w_{i2}, \dots, w_{iN_{w_i}})$ are multiplied with the distance from the producer to the i^{th} injector and summed together. This sum is multiplied by a penalty coefficient c and added to the MSE to obtain the loss function that is minimized during training. This way, long connections are automatically penalized more and the strength of the regularization is set by the penalty coefficient, c in this expression.

2.4.4 Field Application

The proposed methodology was applied to a model cut out from a real field case (Petra Nova) introduced by Olalotiti-Lawal, Onishi, Kim, et al. (2019), which focused on a CCUS project wherein a mature oil field under CO_2 was subjected to water-alternating gas injection.



(a) (b)

Figure 27. Field-scale case model description showing well configuration and heterogeneity (areal and vertical).

The injectors are operated under the rate constrained mode with realistic injection rate schedules, producers operated under stable BHP conditions, and both having intermittent periods of shut-in. A flowchart showing the application of the penalty-based SRU to infer well connectivity for the field-scale case is given in Figure 28.

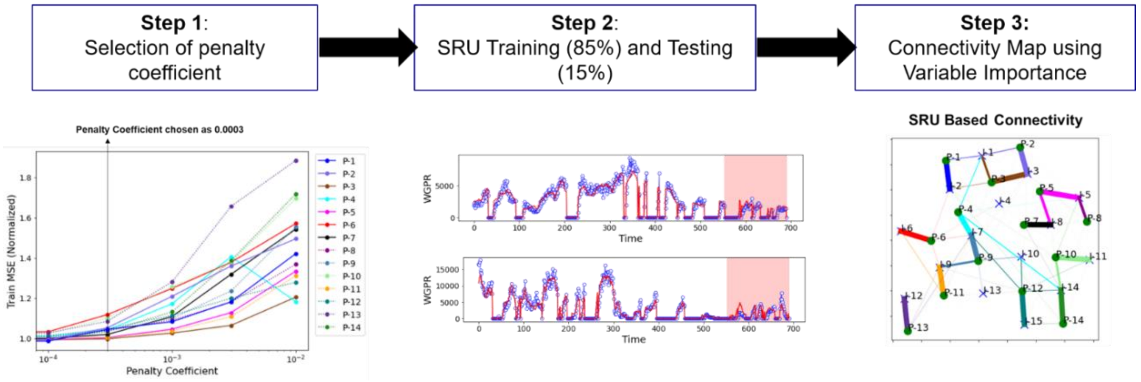
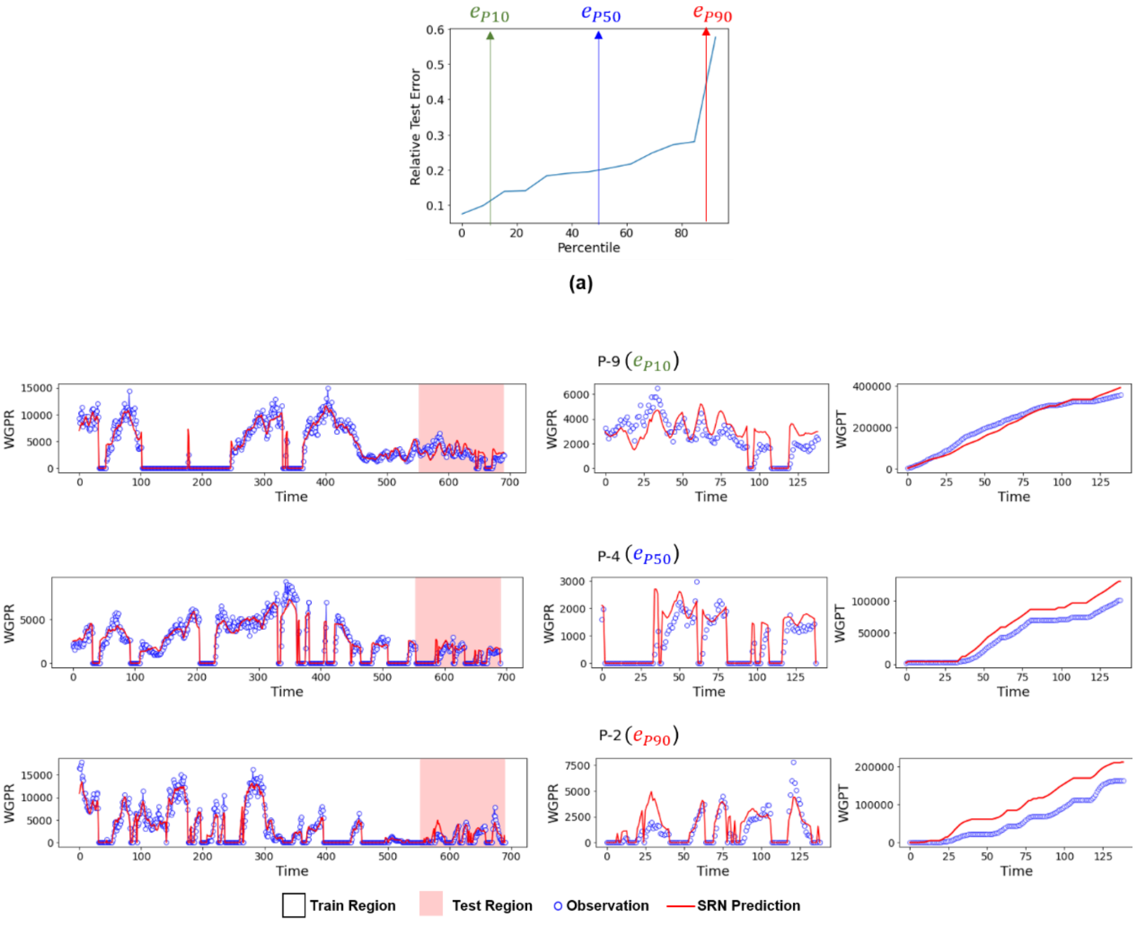


Figure 28. The application of the penalty-function based SRU workflow to the field-scale case. First, a number of SRU models with different penalty coefficients are trained and the best value of penalty coefficient is chosen to be 0. 0003.Secondly, the SRU with the chosen penalty coefficient is trained on 85% of the data and tested on the rest 15% to verify prediction. Finally, we compute the permutation variable importance and generate the connectivity map.

The resulting regression performance is shown in Figure 29. Since we have 14 producers, we only show the wells corresponding to P10, P50 and P90 relative errors in Figure 29 (b). For example, out of all the wells for which we make predictions, producer P-9 exhibits one of the best testing performances, whereas P-4 represents the median prediction performance. Likewise, producer P-2, whose relative test error corresponds to P90 in the error distribution indicates one of the worst regression performances of the SRU in this case.



(a)

(b)

Figure 29. (a) The distribution of the relative test error for all 14 producers, computed as shown in Eq. 8. (b) A selected number of regression fits (corresponding to P90, P50 and P10 errors) are shown. The left panel shows the regression fit for the training and test regions. The middle panel shows the test region zoomed in for better clarity. The right panel shows the cumulative gas production as per observed data and the SRU predictions.

Subsequently, the permutation variable importance was computed for each injector-producer pair. We validate the connectivity map based on SRU variable importance by comparing with the streamline-based connectivity maps (Figure 30). The top panel Figure 30 (a, b) shows all the connections inferred by both methods. The strong fluxes were filtered by setting a threshold on the normalized connectivities and plotted in the bottom panel Figure 30 (c, d). It is seen that the strong connection inferred by the SRU is in almost perfect agreement with those based on streamlines. However, we see that the inferences of weaker fluxes are highly unreliable and indicates the need for a better model, which may be achieved with more data.

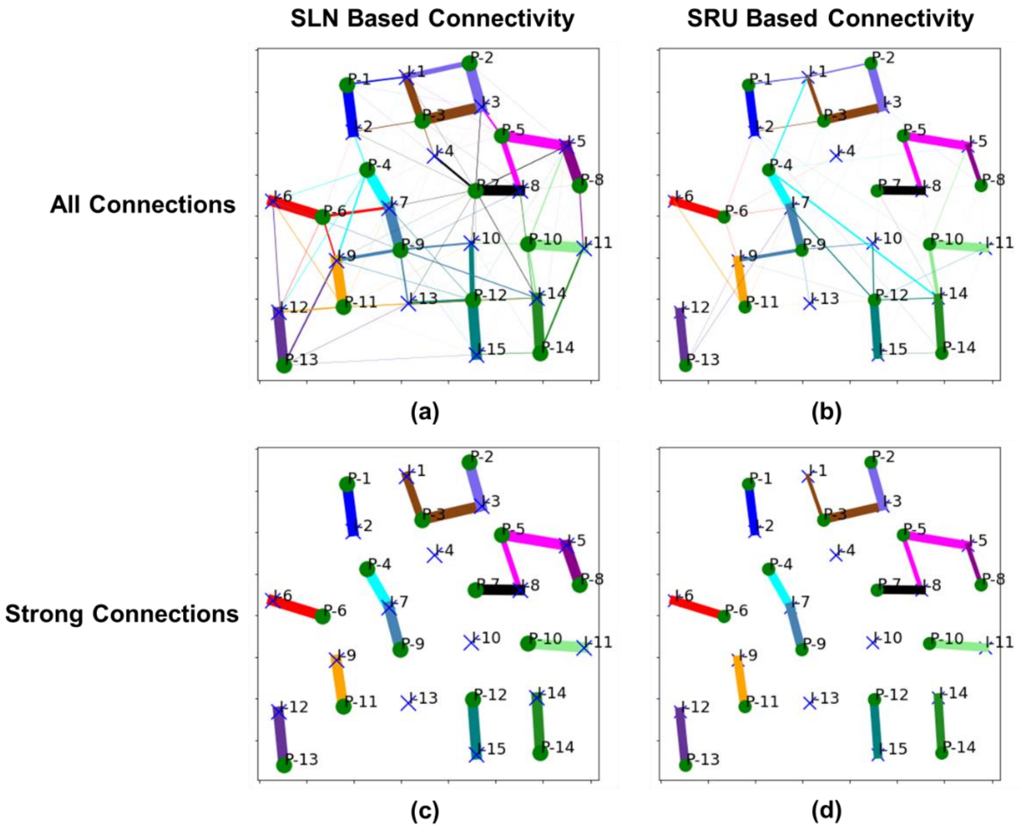


Figure 30. Connectivity maps generated from the proposed methodology compared with average streamline fluxes. The proposed method picks up all strong fluxes even though the inference of weak fluxes is highly uncertain

2.5 Battelle’s Contribution

Texas A&M University, the lead organization in the project, was primarily responsible for the development of tomographic approaches for CO_2 plume mapping in conjunction with distributed pressure, temperature and seismic onset time data. Battelle, as a subcontractor, was primarily responsible for the development of analytical and empirical methods for analyzing transient injection rate and pressure data from point/line sources such as injection and monitoring wells. An additional area of emphasis for Battelle was the use of machine learning for such tasks as inferring reservoir connectivity information from injection-production data, and identifying variable importance for machine learning-based proxy models developed from full-physics simulations. The two organizations also collaborated on the application of the tomographic inversion methodology for a field data set. The contributions of Battelle as part of this project are documented in Appendix-B.

3.0 Summary and Conclusions

The following are the major conclusion for CO2 Plume Tracking Using Pressure and Production Data (section 2.1):

- A two-stage model calibration involving a global update followed by local update is applied to a CO_2 EOR pilot. The global update is directed towards matching field wide reservoir energy and gross fluid movement while the streamline-based local update is used to match the well-wise oil rate, gas rate, water cut and CO_2 production. Historical boundary fluxes from a history matched full field model is imposed at the pilot boundaries to rigorously account for the fluid movement in the field.
- Our model calibration method resulted in an updated model with good predictive capability. The calibrated model was validated by comparing with two sets of independent observation data: the CO_2 recovery and saturation logs at two observation wells. The updated model shows good agreement with the observed data in terms of CO_2 recovery and captures major trends in the reservoir saturation log.
- We further analyzed the implication of the model calibration results using flow diagnostics, particularly the Lorenz curves and tracer sweep efficiencies. The results revealed that hydraulic connectivity is underestimated in the initial model, and are consistent with flow responses at the producers and the CO_2 plume migration.

The following are the major conclusion for Seismic Time-Lapse ‘Onset’ Times for Imaging Front Propagation (section 2.2):

- The sensitivity analysis of the impact of seismic survey frequency shows that a useful onset time map can still be extracted from the infrequent seismic data through interpolation between successive surveys.
- The onset time approach is not effective if there is a change in the dominant underlying physics that is not captured by the seismic surveys, such as

changing from saturation dominated case to pressure and temperature dominated case. The change in the dominant physics will result in the change of seismic response and the interpolation may not give an accurate approximation to the true seismic response.

- The inversion performance of the onset time inversion method was compared with traditional seismic amplitude inversion method. The Brugge benchmark case shows that onset time inversion method obtains comparable model update while being more efficient because of the significant data reduction obtained from converting multiple sets of seismic data into a single onset time map.
- The nonlinearity associated with the onset time inversion and amplitude inversion method was quantitatively investigated. It is found that the nonlinearity in onset time inversion is not only spatially sparser, but also is smaller than the amplitude nonlinearity by several orders of magnitude. The smaller nonlinearity contributes to the superior convergence performance and more robust inversion result of the onset time inversion method.

The following are the major conclusion for Data Assimilation: Using Pressure, Temperature and Production Data for Plume Imaging (section 2.3):

- The thermal tracer TOF and the GTT misfit derived from the DTS data are incorporated into the streamline-based history matching workflow for the first time.
- A hierarchical history matching workflow incorporating GA and the proposed streamline-based DTS data inversion method is applied to a field CO2 Sequestration project in Michigan. The misfits of the distributed pressure measurements and DTS data at both injection well and monitoring well are significantly reduced. The thermal tracer onset time profile and individual temperature curves show great improvements.
- The CO2 plume is imaged by streamlines generated from the history matched model. The visualization shows that the vertical movement of CO2 is restricted, and CO2 stays in the targeted injection zone, which is consistent with an independent warmback analysis of the temperature data.

The following are the major conclusion for Field Application: Identifying Well Connectivities (section 2.4):

- The inputs, outputs and loss function of the SRU model are customized to the CO2 WAG problem. We perform input feature selection to capture underlying patterns with limited data. A penalty function is introduced to include well location information for regularizing the neural net weights.
- Streamlines are used for quantifying physics-based inter-well connectivity for comparison with the SRU model.

- Synthetic and field-scale tests show that the dominant connections identified by the data-driven SRU method and streamline method are in close agreement. Moreover, time-cost for the data-driven method is trivial. This makes the proposed method highly efficient and advantageous for practical field applications.

4.0 References

Abbaszadeh M, and Kamal M, 1989. Pressure-Transient Testing of Water-Injection Wells, SPE Reservoir Engineering, 4 (1),115–124 SPE-16744-PA.

Agada, S., Geiger, S., Elsheikh, A. et al. 2017. Data-driven surrogates for rapid simulation and optimization of WAG injection in fractured carbonate reservoirs. Petroleum Geoscience 23 (2): 270-283.

Achen, Christopher H. 1982. Interpreting and using regression, Vol. 29: Sage.

Albertoni, A., & Lake, L. W. (2003). Inferring interwell connectivity only from well-rate fluctuations in waterfloods. SPE reservoir evaluation & engineering, 6(01), 6-16.

Alhuthali, A.H., Datta-Gupta, A., Yuen, B. et al. 2010. Optimizing smart well controls under geologic uncertainty. Journal of Petroleum Science and Engineering 73 (1): 107-121.

Alhuthali, Ahmed, Oyerinde, Adedayo, and Datta-Gupta, Akhil. 2007. Optimal Waterflood Management Using Rate Control. SPE Reservoir Evaluation & Engineering 10 (05): 539-551. https://doi.org/10.2118/102478-PA.

Ampomah, W., Balch, R.S., Grigg, R.B. et al. 2017. Co-optimization of CO2-EOR and storage processes in mature oil reservoirs. Greenhouse Gases: Science and Technology 7 (1): 128-142.

Arts R, Eiken O, Chadwick A, Zweigel P, van der Meer B, Zinszner B, 2004. Monitoring of CO2 injection at Sleipner using time-lapse seismic data. Energy, 29, 1383–1392.

Azzolina, N.A., Nakles, D.V., Gorecki, C.D. et al. 2015. CO2 storage associated with CO2 enhanced oil recovery: a statistical analysis of historical operations. International Journal of Greenhouse Gas Control 37: 384-397.

Azizi E and Cinar Y, 2013. Approximate analytical solutions for CO2 injectivity in saline formations. SPE Reservoir Evaluation & Engineering.

Bai, S., Song, K.-p., Yang, E.-l. et al. 2011. Optimization of water alternating gas injection parameters of CO2 flooding based on orthogonal experimental design. Special Oil & Gas Reservoirs 1: 032.

Bates, D.M. and Watts, D.G. 1980. Relative Curvature Measures of Nonlinearity. Journal of the Royal Statistical Society. Series B (Methodological): 1-25.

Bao, A., Gildin, E., Huang, J., Coutinho, E. J. R. et al. (2020). Data-Driven End-To-End Production Prediction of Oil Reservoirs by EnKF-Enhanced Recurrent Neural Networks. Proc., SPE Latin American and Caribbean Petroleum Engineering Conference. https://doi.org/10.2118/199005-MS.

Bhark, E., Jafarpour, B., and Datta-Gupta, A. 2011. A generalized grid connectivity–based parameterization for subsurface flow model calibration. Water Resources Research 47 (6).

Bhark, E.W. 2011. Multiscale Spectral-Domain Parameterization for History Matching in

Structured and Unstructured Grid Geometries, Texas A&M University.

Bhark, E.W. and Dehghani, K. 2014. Assisted history matching benchmarking: Design of experiments-based techniques. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Brooks, R.H. and Corey, A.T. 1964. Hydraulic properties of porous media and their relation to drainage design. Trans. ASAE 7 (1): 26-0028.

Bergmann P, Ivandic M, Norden B, Rücker C, Kiessling D, Lüth S, Schmidt-Hattenberger C, and Juhlin C, 2014. Combination of seismic reflection and constrained resistivity inversion with an application to 4D imaging of the CO2 storage site, Ketzin, Germany. Geophysics, 97, 2, B37-B50.

Behrens, R., Condon, P., Haworth, W. 2002. 4D Seismic Monitoring of Water Influx at Bay Marchand: The Practical Use of 4d in an Imperfect World. SPE Reservoir Evaluation & Engineering 5 (5): 410-420. DOI: 10.2118/79961-PA.

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1-4). Springer, Berlin, Heidelberg.

Brauchler R, Cheng JT, Dietrich P, Everett M, Johnson B, Liedl R, Sauter M, 2007. An inversion strategy for hydraulic tomography: coupling travel time and amplitude inversion. J. Hydrol. 345 (3), 184–198.

Brauchler R, Hu R, Dietrich P, Sauter M, 2011. A field assessment of high-resolution aquifer characterization based on hydraulic travel time and hydraulic attenuation tomography. Water Resour. Res. 47 (3), W03503.

Breiman, L. (2001). Random forests. Machine learning. 45 (1): 5-32.

Bring, J. (1996). A geometric approach to compare variables in a regression model. The American Statistician. 50 (1): 57-62.

Budescu, D. V. (1993). Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. Psychological bulletin 114 (3): 542.

Butler JJ, McElwee CD, Bohling GC, 1999. Pumping tests in networks of multilevel sampling wells: motivation and methodology. Water Resour. Res. 35 (11), 3553–3560.

Butsch R, Brown AL, Bryans B, Kolb C, Hovorka S, 2013. Integration of well-based subsurface monitoring technologies: Lessons learned at SECARB study, Cranfield, International Journal of Greenhouse Gas Control, 18, 409-420.

Carrigan C, Yang X, LaBrecque DJ, Larsen D, Freeman D, Ramirez AL, Daily W, Aines R, Newmark R, Friedmann J, Hovorka S, 2013. Electrical resistance tomographic monitoring of CO2 movement in deep geologic reservoirs. International Journal of Greenhouse Gas Control, 18, 401–408.

Chadwick, A., Williams, G., Delepine, N., Clochard, V., Labat, K., Sturton, S., ... & Arts, R. (2010). Quantitative analysis of time-lapse seismic monitoring data at the Sleipner CO2 storage operation. The Leading Edge.

Chang, H., Lo, T.S., Ring, W. et al. 1993. The Effects of Injectant-Enrichment Level on Oil Recovery in Horizontal, Gravity-Tongue-Dominated Enriched-Gas Drives. In Proc., SPE Western Regional Meeting Society of Petroleum Engineers.

Chen, B. and Reynolds, A.C. 2017. Optimal control of ICV's and well operating conditions for the water-alternating-gas injection process. Journal of Petroleum Science and Engineering 149: 623-640.

Chen, B. and Reynolds, A.C., 2018. CO2 water-alternating-gas injection for enhanced oil recovery: Optimal well controls and half-cycle lengths. Computers & Chemical Engineering, 113, pp.44-56.

Chen, S., Li, H., Yang, D. et al. 2010. Optimal parametric design for water-alternating-gas (WAG) process in a CO2-miscible flooding reservoir. Journal of Canadian petroleum technology 49 (10): 75-82.

Cheng, H., Datta-Gupta, A., & He, Z. (2005). A comparison of travel-time and amplitude matching for field-scale production-data integration: Sensitivity, nonlinearity, and practical implications. SPE Journal, 10(01), 75-90.

Chen, Hongquan (2018). Streamline Tracing and Sensitivity Calculation in Fractured Reservoir with Complex Geometry: Field Application to History Matching and Flood Optimization. Doctoral dissertation, Texas A & M University. Available electronically from http://hdl .handle.net/1969.1/174386.

Chen, H., Yang, C., Datta-Gupta, A., Zhang, J., Chen, L., Liu, L., … Bahar, A. (2019, March 22). A Hierarchical Multiscale Framework for History Matching and Optimal Well Placement for a HPHT Fractured Gas Reservoir, Tarim Basin, China. International Petroleum Technology Conference. DOI:10.2523/IPTC-19314-MS

Chen, H., Onishi, T., Olalotiti-Lawal, F., & Datta-Gupta, A. (2020). Streamline tracing and applications in embedded discrete fracture models. Journal of Petroleum Science and Engineering, 106865.

Chen, H., Onishi, T., Park, J. & Datta-Gupta, A. 2021. Computing Pressure Front Propagation Using the Diffusive-Time-of-Flight in Structured and Unstructured Grid Systems via the Fast-Marching Method. SPE Journal: 1-21.

Chen, H., Park, J., Datta-Gupta, A., Shekhar, S., Grover, K., Das, J., ... & Chitale, A. (2020, August). Improving Polymerflood Performance Via Streamline-Based Rate Optimization: Mangala Field, India. In SPE Improved Oil Recovery Conference. Society of Petroleum Engineers.

Chen, H., Yang, C., Datta-Gupta, A., Zhang, J., Chen, L., Liu, L., … Bahar, A. (2020). Fracture Inference and Optimal Well Placement Using a Multiscale History Matching in a HPHT Tight Gas Reservoir, Tarim Basin, China. Upstream Oil and Gas Technology, Volume 2, 2020. Doi：10.1016/j.upstre.2020.100002.

Chen, H., Yao, C., Datta-Gupta, A., et al. (2020, October). Identification of Fractures and Preferential Flow Paths Using Streamlines and Dynamic Data in Dual Porosity Dual Permeability Reservoir Models. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Cui, J., Yang, C., Zhu, D., & Datta-Gupta, A. (2016, December 1). Fracture Diagnosis in Multiple-Stage-Stimulated Horizontal Well by Temperature Measurements With Fast Marching Method. Society of Petroleum Engineers Journal. doi:10.2118/174880-PA.

Curtis, M., Witterholt, E., et al (1973). Use of the temperature log for determining flow rates in producing wells. In Fall Meeting of the Society of Petroleum Engineers of AIME. Society of Petroleum Engineers.

Cheng, H., Wen, X.-H., Milliken, W.J. et al. 2004. Field experiences with assisted and automatic history matching using streamline models. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Crameik, T. and Plassey, J. 1972. Carbon Dioxide Injection Project SACROC Unit, Scurry

County, Texas. In Proc., Annual Meeting Papers, Division of Production American Petroleum Institute.

Chung, J., Gulcehre, C., Cho, K. et al. (2015). Gated feedback recurrent neural networks. In International conference on machine learning (pp. 2067-2075). PMLR.

Cordes, C., & Kinzelbach, W. (1992). Continuous groundwater velocity fields and path lines in linear, bilinear, and trilinear finite elements. Water resources research, 28(11), 2903-2911.

Coulibaly, P. and Baldwin, C.K. (2005). Nonstationary hydrological time series forecasting using nonlinear dynamic methods. Journal of Hydrology 307 (1-4): 164-174.

Daley TM, Myer LR, Peterson JE, Majer EL, Hoversten GM, 2008. Time-lapse crosswell seismic and VSP monitoring of injected CO2 in a brine aquifer. Environmental Geology, 54 (8), 1657-1665.

Daley TM, Freifeld BM, Ajo-Franklin J, Dou S, Pevzner R, Shulakova V, Kashikar S, Miller DE, Goetz J, Henninges J, Lueth S, 2013. Field testing of fiber-optic distributed acoustic sensing (DAS) for subsurface seismic monitoring. The Leading Edge, 32 (6), 699-706.

Datta-Gupta, Akhil, Xie, Jiang, Gupta, Neha et al. 2011. Radius of Investigation and its Generalization to Unconventional Reservoirs. Journal of Petroleum Technology 63 (07): 52-55. https://doi.org/10.2118/0711-0052-JPT.

Datta-Gupta, A. and King, M.J. 2007. Streamline simulation: theory and practice. Society of Petroleum Engineers.

Datta-Gupta A, Yoon SS, Nordaas K and Vasco DW, 2001. Streamlines, Ray Tracing and Production Tomography: Generalization to Compressible Flow. Petroleum Geoscience special issue on Geostatistics, 7, S75-S86.

Dadashpour, M., Ciaurri, D.E., Mukerji, T. 2010. A Derivative-Free Approach for the Estimation of Porosity and Permeability Using Time-Lapse Seismic and Production Data. Journal of Geophysics and Engineering 7 (4): 351-368. DOI: 10.1088/1742-2132/7/4/002.

Dadashpour, M., Echeverría-Ciaurri, D., Kleppe, J. 2009. Porosity and Permeability Estimation by Integration of Production and Time-Lapse near and Far Offset Seismic Data. Journal of Geophysics and Engineering 6 (4): 325.

Dodds K, Krahenbuhl R, Reitz A, Li Y, Hovorka S, 2013. Evaluating time-lapse borehole gravity for CO2 plume detection at SECARB Cranfield. International Journal of Greenhouse Gas Control, 18, 421-429.

Duru, O. O., & Horne, R. N. (2010). Modeling reservoir temperature transients and reservoir-parameter estimation constrained to the model. SPE Reservoir Evaluation & Engineering, 13(06), 873-883.

Duijndam, A. 1988a. Bayesian Estimation in Seismic Inversion. Part I: Principles1. Geophysical Prospecting 36 (8): 878-898.

Duijndam, A. 1988b. Bayesian Estimation in Seismic Inversion. Part II: Uncertainty Analysis1. Geophysical Prospecting 36 (8): 899-918.

Deutsch, C. V., & Journel, A. G. (1992). Geostatistical software library and user’s guide. New York, 119(147).

Duncan, I.J., Nicot, J.-P., and Choi, J.-W. 2008. Risk Assesment for Future CO2 Sequestration Projects Based CO2 Enhanced Oil Recovery in the U.S. Paper presented at the 9th International Conference on Greenhouse Gas Control

Technologies (GHGT-9), 16-20 November, Washington, D.C.

de Holanda, R.W., Gildin, E., & Jensen, J.L. (2018). A generalized framework for Capacitance Resistance Models and a comparison with streamline allocation factors. Journal of Petroleum Science and Engineering. 162: 260-282.

Dunis, C.L. and Huang, X. (2002). Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. Journal of Forecasting. 21 (5): 317-354.

Dutta, K. & Sarma, K.K. (2012). Multiple feature extraction for RNN-based Assamese speech recognition for speech to text conversion application. In 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS) (pp. 600-603). IEEE.

Efendiev, Y., Datta - Gupta, A., Hwang, K., Ma, X., & Mallick, B. (2008). Bayesian partition models for subsurface characterization. Large-Scale Inverse Problems and Quantification of Uncertainty, 107-122.

Ennis-King, J. and Paterson, L., Engineeing Aspects of Geological Sequestration of Carbon Dioxide, SPE 77809, in Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition, 8-10 October 2002, Melbourne, Australia.

Ettehadtavakkol, A., Lake, L.W., and Bryant, S.L. 2014. CO2-EOR and storage design optimization. International Journal of Greenhouse Gas Control 25: 79-92.

Fatemi, E., B. Engquist, and S. Osher, Numerical solution of the high frequency asympoptic expansion for the scalar wave equation, J. Comp. Phys., 120, 145-155, 1995.

Fujita, Y., Datta-Gupta, A., & King, M. J. (2016, December 1). A Comprehensive Reservoir Simulator for Unconventional Reservoirs That Is Based on the Fast Marching Method and Diffusive Time of Flight. Society of Petroleum Engineers Journal. doi:10.2118/173269-PA.

Gasda, S.E. (2004). CO2 Sequestration into a Mature Sedimentary Basin: Determining the Capacity and Leakage Potential of a Saline Aquifer Formation.

Gupta, N., Paul, D., Cumming, L., Place, M., & Mannes, R. G. (2014). Testing for large-scale CO2-enhanced oil recovery and geologic storage in the Midwestern USA. Energy Procedia, 63, 6393-6403.

Ghaderi, S.M., Clarkson, C.R., and Chen, Y. 2012. Optimization of WAG process for coupled CO2 EOR-storage in tight oil formations: an experimental Design approach. In Proc., SPE Canadian Unconventional Resources Conference Society of Petroleum Engineers.

Gassmann, F. 1951. Elastic Waves through a Packing of Spheres. Geophysics 16 (4): 673-685. DOI: 10.1190/1.1437718.

Grimstad, A.-A. and Mannseth, T. 2000. Nonlinearity, Scale, and Sensitivity for Parameter Estimation Problems. SIAM Journal on Scientific Computing 21 (6): 2096-2113.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2008). A novel connectionist system for unconstrained handwriting recognition. IEEE transactions on pattern analysis and machine intelligence, 31(5), 855-868.

Grömping, Ulrike. (2015). Variable importance in regression models. Wiley Interdisciplinary Reviews: Computational Statistics, 7 (2): 137-152.

Ha-Duong, M. and Keith, D.W., 2003. Carbon storage: the economic efficiency of storing CO2 in leaky reservoirs, Clean Techn Environ Policy, 5, 181-189.

Han, L. and Gu, Y. 2014. Optimization of miscible CO2 water-alternating-gas injection in the Bakken formation. Energy & Fuels 28 (11): 6811-6819.

Hastings, W.K. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57 (1): 97–109. doi:10.1093/biomet/57.1.97.

Hawkins, D. M. (2004). The problem of overfitting. Journal of chemical information and computer sciences, 44(1), 1-12.

He, Z., Datta-Gupta, A., and Yoon, S. 2002. Streamline-Based Production Data Integration with Gravity and Changing Field Conditions. SPEJ 7 (4): 423-436. SPE-81208-PA.

He, M., Luis, S., Rita, S. et al. 2011. Risk assessment of CO2 injection processes and storage in carboniferous formations: a review. Journal of Rock Mechanics and Geotechnical Engineering 3(1): 39-56

He, Z., Yoon, S., Datta-Gupta, A., 2002. Streamline-based production data integration with gravity and changing field conditions. SPE Journal 7 (4), 423–436.

Hesse, M.A., and Woods, A.W., Buoyant Dispersal of CO2 during Geologic Storage, Geophysical Research Letters, 37.pL01403 (2010).

Hetz, G., Kim, H., Datta-Gupta, A., King, M. J., Przybysz-Jarnut, J. K., Lopez, J. L., & Vasco, D. (2017, October 9). History Matching of Frequent Seismic Surveys Using Seismic Onset Times at the Peace River Field, Canada. Society of Petroleum Engineers. DOI: 10.2118/187310-MS.

Hetz, G. (2017). Integration of Time Lapse Seismic Data Using Onset Time and Analysis of Spatial Resolution. Doctoral dissertation, Texas A & M University. Available electronically from http://hdl.handle.net /1969.1/169561.

Hetz, G., Datta-Gupta, A., Przybysz-Jarnut, J. K., Lopez, J. L., & Vasco, D. W. (2020). Using onset times from frequent seismic surveys to understand fluid flow at the Peace River Field, Canada. Geophysical Journal International, 223(3), 1610-1629.

Hewson, C.W. and Leeuwenburgh, O., 2017. CO2 water-alternating-gas flooding optimization of the Chigwell Viking I Pool in the western Canadian sedimentary basin. SPE Reservoir Simulation Conference. Society of Petroleum Engineers.

Heffer, K.J., Fox, R.J., McGill, C.A. et al. (1997). Novel Techniques Show Links between Reservoir Flow Directionality, Earth Stress, Fault Structure and Geomechanical Changes in Mature Waterfloods. SPE Journal 2 (02): 91-98. https://doi.org/10.2118/30711-PA.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation 9 (8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Holanda, R.W., Gildin, E., Jensen, J.L. et al. (2018). A state-of-the-art literature review on capacitance resistance models for reservoir characterization and performance forecasting. Energies 11 (12): 3368.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities (in eng). Proceedings of the National Academy of Sciences of the United States of America 79 (8): 2554-2558.

Hosseini, SA, 2014. Time Lapse Compressibility Monitoring for Detection of CO2 Leakage in Brine Aquifers. Energy Procedia, 63, 4459–4463.

Hill, M.C. 2000. Methods and guidelines for effective model calibration. In Building Partnerships, edition, 1-10.

Hu L, Bayer P, Alt-Epping P, Tatomir A, Sauter M, and Brauchler R, 2015. Time-lapse pressure tomography for characterizing CO2 plume evolution in a deep saline aquifer. Intl. J. Greenhouse Gas Control, 39, 91-106.

Ivandic M, Juhlin C, Lüth S, Bergmann P, Kashubin A, Sopher D, Ivanova A, Baumann G, Henninges J, 2015. Geophysical monitoring of CO2 at the Ketzin storage site: new insights into the plume development. International Journal of Greenhouse Gas Control, 32, 90-105.

Jahani, N., Ambía, J., Fossum, K., Alyaev, S., Suter, E., & Torres-Verdín, C. (2021, May). Real-Time Ensemble-Based Well-Log Interpretation for Geosteering. In SPWLA 62nd Annual Logging Symposium. https://doi.org/10.30632/SPWLA-2021-0105. OnePetro

Johns, R., Sah, P., and Solano, R. 2002. Effect of dispersion on local displacement efficiency for multicomponent enriched-gas floods above the minimum miscibility enrichment. SPE Reservoir Evaluation & Engineering 5 (01): 4-10.

Jimenez, E., Datta-Gupta, A., & King, M. J. (2010). Full-field streamline tracing in complex faulted systems with non-neighbor connections. SPE Journal, 15(01), 7-17.

Kam, D. and Datta-Gupta, A., 2016. Streamline-based transport tomography with distributed water arrival times. SPE Reservoir Evaluation & Engineering, 19(02), pp.265-277.

Kang, S. (2014, October). A model segmentation from spectral clustering: new zonation algorithm and application to reservoir history matching. In SPE Annual Technical Conference and Exhibition. OnePetro.

Karhunen, K. 1947. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. Universitat Helsinki.

Kempka T and Kühn M, 2013. Numeric al simulations of CO2 arrival times and reservoir pressure coincide with observations from the Ketzin pilot site, Germany. Environmental Earth Sciences, 70(8), 3675-3685.

Kiessling D, Schmidt-Hattenberger C, Schuett H, Schilling FR, Krueger K, Schoebel B, Danckwardt E, Kummerow J, 2010. Geoelectrical methods for monitoring geological CO2 storage: First results from cross-hole and surface-downhole measurements from the CO2SINK test site at Ketzin (Germany). International Journal of Greenhouse Gas Control, 4, 816–826.

Kokal, S., Sanni, M., and Alhashboul, A. 2016. Design and Implementation of the First CO2-EOR Demonstration Project in Saudi Arabia. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Kumar S, Mandal A. A comprehensive review on chemically enhanced water alternating gas/CO2 (CEWAG) injection for enhanced oil recovery. Journal of Petroleum Science and Engineering. 2017 Aug 1;157:696-715.

Kennett, B. 1983. Seismic Wave Propagation in Stratified Media: Cambridge University Press Original edition. ISBN 0521239338.

Kim, Y.D. & Durlofsky, L.J. (2021). A Recurrent Neural Network–Based Proxy Model for Well-Control Optimization with Nonlinear Output Constraints. SPE Journal: 1-21. https://doi.org/10.2118/203980-PA.

Kruskal, W. (1987). Relative importance by averaging over orderings. The American Statistician 41 (1): 6-10.

Landrø, M., Digranes, P., and Strønen, L. 2001. Mapping Reservoir Pressure and

Saturation Changes Using Seismic Methods-Possibilities and Limitations. First break 19 (12): 671-684.

Lawal, Feyi, Onishi, Tsubasa, Datta-Gupta, Akhil, Fujita, Yusuke Kenji Hagiwara, Post-Combustion CO2 EOR Development in a Mature Field: Model Calibration Using a Hierarchical Approach, SPE 187116 to be presented at the SPE Annual Technical Conference and Exhibition held in San Antonio, Texas, 9-11 October 2017.

Leach, A., Mason, C.F., and van‘t Veld, K. 2011. Co-optimization of enhanced oil recovery and carbon sequestration. Resource and Energy Economics 33 (4): 893-912.

Loeve, M. 1978. Probability theory, vol. ii. Graduate texts in mathematics 46: 0-387.

Lophaven, S.N., Nielsen, H.B., Søndergaard, J. et al. 2002. DACE: a Matlab kriging toolbox, version 2.0. IMM, Informatics and Mathematical Modelling, The Technical University of Denmark.

La Follett, J. R., Wills, P., Lopez, J. L., Przybysz-Jarnut, J. K., & van Lokven, M. 2015. Continuous Seismic Reservoir Monitoring at Peace River: Initial Results and Interpretation. Society of Exploration Geophysicists. DOI: 10.1190/segam2015-5883275.1.

Landrø, M., Digranes, P., and Strønen, L. 2001. Mapping Reservoir Pressure and Saturation Changes Using Seismic Methods-Possibilities and Limitations. First break 19 (12): 671-684.

Landrø, M., Solheim, O.A., Hilde, E. 1999. The Gullfaks 4D Seismic Study. Petroleum Geoscience 5 (3): 213-226.

Li, G., 4D Seismic monitoring of CO2 flood in a thin fractured carbonate reservoir, The leading Edge, July 2003.

Li, Z., Yin, J., Zhu, D, and Datta-Gupta, A., “Using downhole temperature measurement to assist reservoir characterization and optimization,” Journal of Petroleum Science and Engineering, 78 (2), August 2011, Pages 454-63.

Liu S, Yamada M, Collier N, and Sugiyama M, 2013. Change-point detection in time-series data by relative density-ratio estimation. Neural Networks, 43, 72–83.

Liu, T. 2018. Seismic Onset Time Inversion Method for Infrequent Seismic Surveys. Society of Petroleum Engineers. DOI: 10.2118/194044-STU.

Liu, T., Chen, H., Hetz, G., & Datta-Gupta, A. (2020). Integration of Time-lapse Seismic Data Using the Onset Time Approach: The Impact of Seismic Survey Frequency. Journal of Petroleum Science and Engineering, 106989.

Liu, W., Liu, W., & Gu, J. (2019). Reservoir Inter-Well Connectivity Analysis Based on a Data Driven Method. In Abu Dhabi International Petroleum Exhibition & Conference. https://doi.org/10.2118/197654-MS.

Lopez, J., Wills, P., La Follett, J. et al. 2015. Real-Time Seismic Surveillance of Thermal EOR at Peace River. In SPE Canada Heavy Oil Technical Conference: Society of Petroleum Engineers. ISBN 1613994028.

Lumley, D., 2010, 4D seismic monitoring of CO2 sequestration: The Leading Edge, 29, 150–155.

Luo, Y., and G. Schuster, 1991a, Wave equation traveltime inversion, Geophysics, 56, 645–653, DOI:10.1190/1.1443081.

Luo, Y., and G. Schuster, 1991b, Wave equation inversion of skeletonized geophysical data: Geophysical Journal International, 105, 289–294, DOI:10.1111/j.1365-246X.1991.tb06713.x.

Ma, X., Al-Harbi, M., Datta-Gupta, A., & Efendiev, Y. (2008). An efficient two-stage sampling method for uncertainty quantification in history matching geological models. SPE Journal, 13(01), 77-87.

Maucec, M., Douma, S. G., Hohl, D., Leguijt, J., Jimenez, E., & Datta-Gupta, A. (2007, January). Streamline-Based History Matching and Uncertainty--Markov-Chain Monte Carlo Study of an Offshore Turbidite Oil Field. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Mavko, G., Mukerji, T., and Dvorkin, J. 2009. The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media: Cambridge university press. Original edition. ISBN 0521861365.

Mawalkar S., Brock D., Burchwell A., Kelley M et al., Where is that CO2 flowing? Using Distributed Temperature Sensing (DTS) technology for monitoring injection of CO2 into a depleted oil reservoir, International Journal of Greenhouse Gas Control, Volume 85, 132-142, 2019

Mathieson, A., Wright, I., Roberts, D. and Ringrose, P., Satellite Imaging to Monitor CO2 Movement at Krechba, Algeria, Energy Proceedia 1 2201-2209, 2009.

McKay, M.D., Beckman, R.J., and Conover, W.J. 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2): 239-245.

Melzer, L.S. 2006. Stranded Oil in the Residual Oil Zone. Melzer Consulting prepared for Advanced Resources International and the US Department of Energy: Office of Fossil Energy0 Office of Oil and Natural Gas.

Melzer, L.S. 2012. Carbon dioxide enhanced oil recovery (CO2 EOR): Factors involved in adding carbon capture, utilization and storage (CCUS) to enhanced oil recovery. Center for Climate and Energy Solutions.

Merchant, D.H. 2010. Life Beyond 80: A look at conventional WAG recovery beyond 80% HCPV injected in CO2 tertiary floods. In Proc., SPE International Conference on CO2 Capture, Storage, and Utilization Society of Petroleum Engineers.

Mindlin, R. 1949. Compliance of Elastic Bodies in Contact. J. of Appl. Mech. 16.

Mindlin, R.D. 1949. Compliance of Elastic Bodies in Contact. Journal of Applied Mechanics 16: 259-268.

Mishra S., M. Kelley, A. Haagsma, P. Ravi Ganesh, J. Main, V. Smith, A. Pasumarti, S. Raziperchikolaee, A. Burchwell, (2020). Numerical modeling of primary production and CO2 injection performance in depleted oil fields in the Michigan Northern Pinnacle Reservoir Trend, Final Topical Report, Midwest Regional Carbon Sequestration Partnership, Columbus, OH.

Mishra S, Kelley M, Zeller E, Slee N, Gupta N, Bhattacharya I and Hammond M, 2012. “Maximizing the value of pressure monitoring data from CO2 sequestration projects,” Proc., International Conference on Greenhouse Gas Technologies (GHGT-11), Kyoto, Japan.

Mishra S, Oruganti YD, Gupta N, Ganesh PR, McNeil C, Bhattacharya I, and Spitznogle G, 2014. Modeling CO2 plume migration based on calibration of injection and post‐injection pressure response at the AEP Mountaineer Project. Greenhouse Gases: Science and Technology, 4(3), 331-356.

Mondal, A., Efendiev, Y., Mallick, B., & Datta-Gupta, A. (2010). Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump

Markov chain Monte Carlo methods. Advances in Water Resources, 33(3), 241-256.

Møyner, O., Krogstad, S., and Lie, K.-A. 2015. The application of flow diagnostics for reservoir management. SPE Journal 20 (02): 306-323.

Nolet G. 1987. Seismic wave propagation and seismic tomography, Nolet, G. (ed), Seismic Tomography. Dordrecht: D. Reidel, 1-23

Nooner, S., Eiken, O., Hermanrud, C., Sasagawa, G., Stenvold, T., and Zumberge, M., Constraints on the in-situ density of CO2 within the Utsira formation from time-lapse seafloor gravity measurements, International Journal of Greenhouse Gas Control, Volume 1, Issue 2, April 2007.

Olalotiti-Lawal, F. 2018. Effective Reservoir Mangement for Carbon Utilization and Storage Applications. Doctoral Dissertation, Texas A&M University.

Olalotiti-Lawal, F. and Datta-Gupta, A., 2019. Multi-resolution Grid Connectivity-Based Reparameterization for Effective Subsurface Model Calibration. Transport in Porous Media, pp.1-29.

Olalotiti-Lawal, F., Onishi, T., Kim, H., Datta-Gupta, A., Fujita, Y., & Hagiwara, K. (2019). Post-combustion carbon dioxide enhanced-oil-recovery development in a mature oil field: model calibration using a hierarchical approach. SPE Reservoir Evaluation & Engineering, 22(03), 998-1014.

Olalotiti-Lawal, F., Onishi, T., Datta-Gupta, A. et al. 2017. Post-Combustion CO2 EOR Development in a Mature Oil Field: Model Calibration Using a Hierarchical Approach. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Onishi, T., Nguyen, M.C., Carey, J.W., Will, B., Zaluski, W., Bowen, D.W., Devault, B.C., Duguid, A., Zhou, Q., Fairweather, S.H. and Spangler, L.H., 2019. Potential CO2 and brine leakage through wellbore pathways for geologic CO2 sequestration using the National Risk Assessment Partnership tools: Application to the Big Sky Regional Partnership. International Journal of Greenhouse Gas Control, 81, pp.44-65.

Onwunalu, J.E. and Durlofsky, L. 2011. A new well-pattern-optimization procedure for large-scale field development. SPE Journal 16 (03): 594-607.

Orr, F.M. 2007. Theory of gas injection processes. Tie-Line Publications.

Osako, I. and Datta-Gupta, A. 2007. A compositional streamline formulation with compressibility effects. In Proc., SPE Reservoir Simulation Symposium Society of Petroleum Engineers.

Oliva, J.B., Póczos, B., & Schneider, J. (2017). The Statistical Recurrent Unit. In International Conference on Machine Learning, (pp. 2671-2680). PMLR.

Quintero, L. F., Manrique, E. J., Linares, J. A., Rarnones, M. A., Natera, J. A., et al (1993). Permeability measurement through temperature logging. In SPWLA 34th Annual Logging Symposium. Society of Petrophysicists and Well-Log Analysts.

Paige, C.C. and Saunders, M.A. 1982. LSQR: An algorithm for sparse linear equations and sparse least squares. ACM transactions on mathematical software 8 (1): 43-71.

Parker, R.L. 1994. Geophysical inverse theory. Princeton university press.

Passone, S. and McRae, G.J. 2007. Probabilistic Field Development in Presence of Uncertainty. In Proc., International Petroleum Technology Conference International Petroleum Technology Conference.

Pham, V. and Halland, E. 2017. Perspective of CO2 for Storage and Enhanced Oil Recovery (EOR) in Norwegian North Sea.

Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R., Chen, Y., Dunlop, N., Vossepoel, F. and Xu, R. 2010. Results of the Brugge benchmark study for flooding optimization and history matching. SPE Reservoir Evaluation & Engineering 13 (3): 391-405. SPE 119094-PA. DOI: 10.2118/119094-PA.

Pruess, K., Xu, T., Apps, J., & Garcia, J. (2001, January). Numerical modeling of aquifer disposal of CO2. In SPE/EPA/DOE Exploration and Production Environmental Conference. Society of Petroleum Engineers.

Pruess K, Oldenburg C, and Moridis G, 2011. TOUGH2 user's guide, version 2.0 (revised): Lawrence Berkeley Laboratory Report LBL-43134, Berkeley, California

Przybysz-Jarnut, J. K., Didraga, C., Potters, J. H. H. M., Lopez, J. L., La Follett, J. R., Wills, P. B., … Brouwer, D. R. 2015. Value of Information of Frequent Time-Lapse Seismic for Thermal EOR Monitoring at Peace River. Society of Petroleum Engineers. DOI:10.2118/175046-MS.

Pan, Y., Deng, L., Zhou, P., & Lee, W. J. (2021). Laplacian Echo-State Networks for production analysis and forecasting in unconventional reservoirs. Journal of Petroleum Science and Engineering, 109068.

Pollock, D. W. (1988). Semianalytical computation of path lines for finite-difference models. Groundwater, 26(6), 743-750.

Prevost, M., Edwards, M. G., & Blunt, M. J. (2002). Streamline tracing on curvilinear structured and unstructured grids. SPE Journal, 7(02), 139-148.

Ramey HJ Jr, 1970. Approximate solutions for unsteady liquid flow in composite reservoirs. J Cdn Pet Tech, 32-37.

Ramirez AL, Newmark RL, Daily WD, 2003. Monitoring carbon dioxide floods using electrical resistance tomography (ERT): Sensitivity studies. Journal of Environmental and Engineering Geophysics, 8, 187–208.

Rasmussen, A. F. (2010, September). Streamline tracing on irregular geometries. In ECMOR XII-12th European Conference on the Mathematics of Oil Recovery. (pp. cp-163). European Association of Geoscientists & Engineers.

Rey, A., Bhark, E., Gao, K. 2012. Streamline-Based Integration of Time-Lapse Seismic and Production Data into Petroleum Reservoir Models. Geophysics 77 (6): M73-M87. DOI: 10.1190/Geo2011-0346.1.

Rey, A., Taware, S.V., Datta-Gupta A., 2010. Use of Time-Lapse Seismic Data for Heterogeneity Characterization During CO2 Sequestration in Saline Aquifers, presented at the SPE International Conference on CO2 Capture, Storage and Utilization, New Orleans, Louisiana, USA, November, 2010

Ren, B. & Duncan, I.J. (2019). Reservoir simulation of carbon storage associated with CO2 EOR in residual oil zones, San Andres formation of West Texas, Permian Basin, USA. Energy 167: 391-401.

Ribeiro, P. M., & Horne, R. N. (2013, September). Pressure and temperature transient analysis: hydraulic fractured well application. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Rossen, W.R., Van Duijn, C., Nguyen, Q.P. et al. 2010. Injection strategies to overcome gravity segregation in simultaneous gas and water injection into homogeneous reservoirs. SPE Journal 15 (01): 76-90.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. Nature 323 (6088): 533-536.

Sayarpour, Morteza. (2008). Development and application of capacitance-resistive models to water/CO_2 floods.

Sarma, P., Durlofsky, L.J., and Aziz, K. 2005. Efficient Closed-Loop Production Optimization under Uncertainty (SPE94241). In Proc., 67th EAGE Conference & Exhibition

Sarma, P., Durlofsky, L.J., and Aziz, K. 2008. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. Mathematical Geosciences 40 (1): 3-32.

Satman A, 1985. An analytical study of interference in composite reservoirs. Soc Pet Eng J, 281-90.

Sato K, Mito S, Horie T, Ohkuma H, Saito H, Watanabe J, and Yoshimura T, 2011. Monitoring and simulation studies for assessing macro-and meso-scale migration of CO 2 sequestered in an onshore aquifer: experiences from the Nagaoka pilot site, Japan. International Journal of Greenhouse Gas Control, 5(1), 125-137.

Schmidt-Hattenberger C, Bergmann P, Labitzke T, Schröder S, Krüger K, Rücker C, Schütt H, 2012. A Modular Geoelectrical Monitoring System as Part of the Surveillance Concept in CO2 Storage Projects. Energy Procedia, 23, 400–407.

Schlumberger. 2014. Technical Description. ECLIPSE Industry-Reference Reservoir Simulator. Schlumberger.

Schlumberger. (2018). ECLIPSE Technical Description Manual. Thermal CO2STORE section.

Sethian JA, 1999. Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science Vol. 3: Cambridge university press.

Shahvali, M., Mallison, B., Wei, K. et al. 2012. An alternative to streamlines for flow diagnostics on structured and unstructured grids. SPE Journal 17 (03): 768-778.

Sharma, M., Taware, S., and Datta-Gupta, A. 2016. Optimizing CO2 Floods Using Rate Control with Smart Wells Under Geologic Uncertainty. In Proc., Abu Dhabi International Petroleum Exhibition & Conference Society of Petroleum Engineers.

Shi J and Malik J, 2000. Normalized Cuts and Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (8): 888-905.

Shook, G.M. and Mitchell, K.M. 2009. A robust measure of heterogeneity for ranking earth models: The F PHI curve and dynamic Lorenz coefficient. In Proc., SPE annual technical conference and exhibition Society of Petroleum Engineers.

Singh V, Cavanagh A, Hansen H, Nazarina B, Idling M and Ringrose P, 2010. “Reservoir modeling of CO2 plume behavior calibrated against monitoring data from Sleipner, Norway,” paper SPE 134891 presented at SPE Annual Technical Conference and Exhibition, Florence, Italy, Sept. 19-22

Solano, R., Johns, R., and Lake, L. 2000. Impact of reservoir mixing on recovery in enriched-gas drives above the minimum miscibility enrichment. In Proc., SPE/DOE Improved Oil Recovery Symposium Society of Petroleum Engineers.

Song, Z., Li, Z., Wei, M. et al. 2014. Sensitivity analysis of water-alternating-CO2 flooding for enhanced oil recovery in high water cut oil reservoirs. Computers & Fluids 99: 93-103.

Stalkup, F. 1998. Predicting the effect of continued gas enrichment above the MME on oil recovery in enriched hydrocarbon gas floods. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Storn, R. and Price, K. 1997. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. J. of Global Optimization, 11: 341–359.

Stone, H. 1973. Estimation of three-phase relative permeability and residual oil data. J. Pet. Technol.;(United States) 12 (4).

Sen, D., Chen, H., Datta-Gupta, A., Kwon, J., & Mishra, S. (2020, October). Data-Driven Rate Optimization Under Geologic Uncertainty. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers. https://doi.org/10.2118/201325-MS.

Sen, D., Chen, H., Datta-Gupta, A., Kwon, J., & Mishra, S. (2021). Machine Learning based Rate Optimization Under Geologic Uncertainty. Journal of Petroleum Science and Engineering. https://doi.org/10.1016/j.petrol.2021.109116.

Sen, D., Ong, C., Kainkaryam, S., & Sharma, A. (2020). Automatic Detection of Anomalous Density Measurements due to Wellbore Cave-in. Petrophysics, 61(05), 434-449. https://doi.org/10.30632/PJV61N5-2020a3.

Sen, D., Raihan, D., & Chidambaram, M. (2014). Multiway continuous hidden Markov model ‐ based approach for fault detection and diagnosis. AIChE Journal, 60(6), 2035-2047.Shadravan, A., Ghasemi, M., and Alfi, M. 2015. Zonal Isolation in Geothermal Wells. Presented at the Fortieth Workshop on Geothermal Reservoir Engineering, 26-28 January, Stanford University, Stanford, California.

Somogyvári, M. and Bayer, P., 2017. Field validation of thermal tracer tomography for reconstruction of aquifer heterogeneity. Water Resources Research, 53(6), pp.5070-5084.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404: 132306.

Suarsana, I. P., & Badril, A. (2011, May). Comparison of Tracer Test Result and Analysis of Connectivity Injector and Producer during Pilot Waterflood Kenali Asam Zone P/1050. In SPE EUROPEC/EAGE Annual Conference and Exhibition. OnePetro.

T. LaForce*, J. Ennis-King, L. Paterson, 2013, Magnitude and duration of temperature changes in geological storage of carbon dioxide. Energy Procedia 37 (2013) 4465 – 4472

Taber, J., Martin, F., and Seright, R. 1997a. EOR screening criteria revisited—part 2: applications and impact of oil prices. SPE Reservoir Engineering 12 (03): 199-206.

Taber, J.J., Martin, F., and Seright, R. 1997b. EOR screening criteria revisited-Part 1: Introduction to screening criteria and enhanced recovery field projects. SPE Reservoir Engineering 12 (03): 189-198.

Tanaka, S., Datta-Gupta, A., and King, M.J. 2014. Compositional Streamline Simulation of CO2 Injection Accounting for Gravity and Capillary Effects Using Orthogonal Projection. In Proc., SPE Improved Oil Recovery Symposium Society of Petroleum Engineers.

Tanaka, S., Kam, D., Datta-Gupta, A. et al. 2015. Streamline-Based History Matching of Arrival Times and Bottomhole Pressure Data for Multicomponent Compositional Systems. In Proc., SPE Annual Technical Conference and Exhibition Society of

Petroleum Engineers.

Tanaka, S., Kam, D., Xie, J. et al. 2017. A Generalized Derivative-Free Rate Allocation Optimization for Water and Gas Flooding Using Streamline-Based Method. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Tang, J., Ehlig-Economides, C., Li, J., Guo, X., Wu, K., Mu, L., Tang, M., Du, X. 2018. Investigation of Multiple Hydraulic Fracture Propagation for Low-pressure Tight Sandstone Reservoirs in Northern Ordos Basin of Changqing Oilfield, China. In: The SPE International Hydraulic Fracturing Technology Conference and Exhibition, Muscat, Oman. SPE-191454-MS. DOI: 10.2118/191454-18IHFT-MS.

Ter Braak, C.J.F. 2006. A Markov Chain Monte Carlo Version of the Genetic Algorithm Differential Evolution: Easy Bayesian Computing for Real Parameter Spaces. Stat Comput., 16: 239-249. doi: 10.1007/s11222-006-8769-1.

Tian, C., & Horne, R. N. (2015, September). Machine learning applied to multiwell test analysis and flow rate reconstruction. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Tian, C., & Horne, R. N. (2016, September). Inferring interwell connectivity using production data. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Tian, C., & Horne, R. N. (2017, October). Recurrent neural networks for permanent downhole gauge data analysis. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Tian, C., & Horne, R. N. (2019). Applying machine-learning techniques to interpret flow-rate, pressure, and temperature data from permanent downhole gauges. SPE Reservoir Evaluation & Engineering, 22(02), 386-401.

Todd, M. R., & Longstaff, W. J. (1972). The development, testing, and application of a numerical simulator for predicting miscible flood performance. Journal of Petroleum Technology, 24(07), 874-882.

Tsang, C.-F., Benson, S.M., Kobelski, B., and Smith, R.E., 2002. Scientific considerations related to regulation development for CO2 sequestration in brine formations, Environmental Geology, 42,275-281.

Vasco, D., Seongsik, Y., and Datta-Gupta, A. 1999. Integrating Dynamic Data Into High-Resolution Reservoir Models Using Streamline-Based Analytic Sensitivity Coefficients. SPE Journal 4 (04): 389-399.

Vasco, D.W., Daley, T.M., and Bakulin, A. 2014. Utilizing the Onset of Time-Lapse Changes: A Robust Basis for Reservoir Monitoring and Characterization. Geophysical Journal International 197 (1): 542-556. DOI: 10.1093/gji/ggt526.

Vasco, D.W., Bakulin, A., Baek, H. 2015. Reservoir Characterization Based Upon the Onset of Time-Lapse Amplitude Changes. Geophysics 80 (1): M1-M14. DOI: 10.1190/Geo2014-0076.1.

Vasco, D.W. and Datta-Gupta, A. 2016. Subsurface Fluid Flow and Imaging: With Applications for Hydrology, Reservoir Engineering, and Geophysics: Cambridge University Press. Original edition. ISBN 1316577945.

Vasco, D.W. and Datta-Gupta, A., 1999. Asymptotic solutions for solute transport: A formalism for tracer tomography. Water Resources Research, 35(1), pp.1-16.

Veettil, D. R. A., & Clark, K. (2020). Bayesian geosteering using sequential monte carlo methods. Petrophysics, 61(01), 99-111.

Villesca, J., Glasbergen, G., Attaway, D. J., et al (2011). Measuring fluid placement of sand consolidation treatments using dts. In SPE European Formation Damage Conference. Society of Petroleum Engineers.

Wang, X., Bussear, T. R., et al (2011). Real time horizontal well monitoring using distributed temperature sensing (dts) technology. In OTC Brasil. Offshore Technology Conference.

Wang, Z. (2012). The Uses of Distributed Temperature Survey (DTS) Data. PhD thesis, Stanford University.

Wallace, M. and Kuuskraa, V. 2014. Near-Term Projections of CO2 Utilization for Enhanced Oil Recovery. National Energy Technology Laboratory.

Watanabe, S., Han, J., Hetz, G. et al. 2017. Streamline-Based Time-Lapse-Seismic-Data Integration Incorporating Pressure and Saturation Effects. SPEJ Society of Petroleum Engineers.

White D, 2009. Monitoring CO2 storage during EOR at the Weyburn-Midale Field. The Leading Edge, 28, 7, 838-842.

White MD, Bacon DH, McGrail BP, Watson DJ, White SK, and Zhang Z, 2012. STOMP Subsurface Transport Over Multiple Phases: STOMP-CO2 and STOMP-CO2e Guide: Version 1.0: Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

Wilson, M. and Monea, M. 2004. IEA GHG Weyburn CO2 monitoring & storage project. Summary report 2000-2004.

Watanabe, S., Han, J., Hetz, G. 2017. Streamline-Based Time-Lapse Seismic Data Integration Incorporating Pressure and Saturation Effects. SPEJ Society of Petroleum Engineers.

Xu, B. and Fahim F., "Uncertainty Reduction in Reservoir Geostatistical Description Using Distributed Temperature Sensing (DTS) Systems Data."Paper presented at the SPE Western Regional Meeting, Anchorage, Alaska, USA, May 2016.

Xue Z, Tanase D, Watanabe J, 2006. Estimation of CO2 saturation from time-lapse CO2 well logging in an onshore aquifer, Nagaoka, Japan. Explor. Geophys., 37, 19–29.

Yamaguchi, S., Ohga, K., Fijioka, M., and Nako, M., Field Teas and History Matching of the CO2 Sequestration Project in Coal Seams in Japan, Int, J. Soc. Mater. Eng. Resour., Vol 13, No. 2, Mar 2006.

Yeh T, Liu S, 2000. Hydraulic tomography: development of a new aquifer test method. Water Resour. Res., 36 (8), 2095–2105.

Yin, J., Park, H.-Y., Datta-Gupta, A. et al. 2011. A hierarchical streamline-assisted history matching approach with global and local parameter updates. Journal of Petroleum Science and Engineering 80 (1): 116-130.

Yousef, A. A., Gentil, P. H., Jensen, J. L., & Lake, L. W. (2006). A capacitance model to infer interwell connectivity from production and injection rate fluctuations. SPE Reservoir Evaluation & Engineering, 9(06), 630-646.

Zapata, V.J. and Lake, L.W. 1981. A theoretical analysis of viscous crossflow. In Proc., SPE Annual Technical Conference and Exhibition Society of Petroleum Engineers.

Zhang, K., Tang, M., Yong, S. S., & Shan, Y. (2016, August). Evaluation of Stage Contribution and Interwell Connectivity During Initial Flowback and Oil Production in a Tight Oil Horizontal Stimulation Using Tracer Technology. In SPE

Asia Pacific Hydraulic Fracturing Conference. Society of Petroleum Engineers.

Zhang, Q., Zhou, A. and Jin, Y. 2008. RM-MEDA: A Regularity Model-Based Multiobjective Estimation of Distribution Algorithm. IEEE Transactions on Evolutionary Computation. 12(1):41–63.

Zhang R, Ghosh R, Sen M, and Srinivasan S, 2013. Time-lapse surface seismic inversion with thin bed resolution for monitoring CO2 sequestration: A case study from Cranfield, Mississippi. International Journal of Greenhouse Gas Control, 18, 430–438

Zhang, Y., Bansal, N., Fujita, Y., Datta-Gupta, A., King, M. J., and Sankaran, S., 2016 From Streamlines to Fast Marching: Rapid Simulation and Performance Assessment of Shale Gas Reservoirs Using Diffusive Time of Flight as a Spatial Coordinate, SPE Journal doi:10.2118/168997-PA.

Zhang, Y., King, M. J., & Datta-Gupta, A. (2012). Robust streamline tracing using inter-cell fluxes in locally refined and unstructured grids. Water Resources Research, 48(6).

Zhou, D., Yan, M., and Calvin, W.M. 2012. Optimization of a mature CO2 flood-from continuous injection to WAG. In Proc., SPE Improved Oil Recovery Symposium Society of Petroleum Engineers.

Zhou, S., Huang, H. & Liu, Y. Pet. Sci. (2008) 5: 87. https://doi.org/10.1007/s12182-008-0015-3

Zhou, C., W. Cai, Y. Luo, G. Schuster, and S. Hassanzadeh, 1995, Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data: Geophysics, 60, 765–773, DOI:10.1190/1.1443815.

Zhou, P., Pan, Y., Sang, H., & Lee, W. J. (2018, September). Criteria for proper production decline models and algorithm for decline curve parameter inference. In Unconventional Resources Technology Conference, Houston, Texas, 23-25 July 2018 (pp. 3535-3551). Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers.

Zuber, V., & Strimmer, K. (2011). High-dimensional regression and variable selection using CAR scores. Statistical Applications in Genetics and Molecular Biology, 10(1).

Zuo, L., Lim, J., Chen, R., Chen, H., Shamekhi, A., & King, M. J. (2021). Continuous streamline trajectories on complex grids. Computational Geosciences, 1-25.

Appendix A. Abstracts for Peer-Reviewed Publications Created by the Project

Paper 1

Feyi Olalotiti-Lawal, Tsubasa Onishi, Akhil Datta-Gupta, Yusuke Fujita, Daiki Watanabe, Kenji Hagiwara, *Model calibration and optimization of a post-combustion CO2 WAG pilot in a mature oil field*, **Fuel**, Volume 255, 2019, 115810, ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2019.115810.

Carbon Dioxide Enhanced Oil Recovery (CO_2 EOR) technology is a proven and mature technology that has been profitably adopted in rejuvenating and sustaining mature fields for almost half a century (Wallace and Kuuskraa, 2014). CO_2 is injected into an oil-bearing formation during which it contacts, mixes with and mobilizes in situ residual oil. Beginning from early 1972 when the CO_2 EOR technology was first commercially deployed in the SACROC unit in the Scurry county of Texas (Crameik and Plassey, 1972), the oil and gas industry has witnessed a widespread adoption of the technology for enhancing the value of mature assets, especially in North America. In 2012, an average of 282,000 bbl/D of oil was produced by CO_2 EOR in the United States and the number is projected to rise over 600,000 bbb/D by the year 2020 (Wallace and Kuuskraa, 2014). Although most applications have been in depleted oil reservoirs in the onshore environment including sandstone, carbonate (Wilson and Monea, 2004) and recently in unconventional tight formations (Ghaderi et al., 2012; Han and Gu, 2014), offshore application of the technology has also started drawing attention (Pham and Halland, 2017). Recently, chemically enhanced water alternating gas/CO_2 method has received attention from the oil and gas industry and the studies are well summarized in Kumar and Mandal, (2017). In addition, Residual Oil Zones (ROZs) are beginning to generate interests due to its potentially substantial value creation through CO_2 EOR (Melzer, 2006).

The CO_2 EOR technology has been applied in different forms, based on specific nature and conditions of the oil-bearing formations for optimal oil recovery. These include continuous CO_2 injection, Water-Alternating-Gas (WAG) Injection, gravity drainage, huff-and-puff, and so on. Regardless of the process employed for CO_2 EOR, proper reservoir management plays a key role in maximizing the value of the asset using this technology (Wallace and Kuuskraa, 2014). This includes reservoir characterization to understand the subsurface flow mechanisms, followed by generation of optimal strategies for improved conformance. Measures to combat flow conformance problems become particularly vital for gas injection processes like CO_2 WAG. This is because of both viscous fingering and gravity override effects, due to differences in multiphase fluid properties including viscosity and density (Orr, 2007). Mitigating conformance issues typically starts with precautionary screening of potential reservoirs (Taber et al., 1997a; Taber et al., 1997b). Development of selected reservoirs then requires effective field management procedures for optimal operational strategies, well placements as well as fluid rates, taking into cognizance subsurface geological features and existing flow dynamics (Møyner et al., 2015; Onwunalu and Durlofsky, 2011; Passone and McRae, 2007; Sarma et al., 2005; Zhou et al., 2012).

This paper is organized as follows: First we provide a background and description of the field and we also discuss the initialization of the pilot model. Second, we report the details of the pilot model calibration, starting from the model parameterization and the optimization strategy adopted to reduce data misfits in the global step of the hierarchical workflow. Next, we describe the local model update algorithm which utilizes streamline-derived arrival time sensitivities to integrate CO_2 breakthrough times into the pilot model. A discussion of the results and validation of the calibrated model using auxiliary observed data is presented next.

Paper 2

Tian Liu, Hongquan Chen, Gill Hetz, Akhil Datta-Gupta, *Integration of time-lapse seismic data using the onset time approach: The impact of seismic survey frequency*, **Journal of Petroleum Science and Engineering**, Volume 189, 2020, 106989, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2020.106989.

The use of seismic onset times has shown great promise for integrating near-continuous seismic surveys for updating geologic models. However, due to the high cost of seismic surveys, such frequent seismic surveys are not commonly available. In this study, we focus on analyzing the impact of seismic survey frequency on the onset time approach aiming to extend the application of onset time to infrequent seismic surveys. In addition, we quantitatively examine the nonlinearity of the onset time method and compare it to the commonly used amplitude inversion method.

We carry out a sensitivity analysis of seismic survey frequency based on the complete seismic survey data (over 175 surveys) of steam injection in a heavy oil reservoir (Peace River Unit) in Canada. Data sets of different survey frequencies are generated by sampling at various time intervals from the complete data sets. Onset time maps based on different survey frequencies are calculated. Our results show that an adequate onset time map can be obtained from the infrequent seismic surveys by interpolation between seismic surveys as long as there is no change in the dominant underlying physics between the successive surveys.

In terms of robustness of the inversion methods, nonlinearity of the onset time method can be smaller than that of the amplitude inversion method by several orders of magnitude. Application to the Brugge benchmark case shows that the onset time method obtains comparable permeability update as the traditional seismic amplitude inversion method with faster computation and improved convergence characteristics.

Paper 3

Yao, Changqing , Chen, Hongquan , Onishi, Tsubasa , Datta-Gupta, Akhil , Mawalkar, Sanjay , Mishra, Srikanta , and Ashwin Pasumarti. *Robust CO2 Plume Imaging Using Joint Tomographic Inversion of Distributed Pressure and Temperature Measurements*. Paper presented at the SPE Annual Technical Conference and Exhibition, Dubai, UAE, September 2021. doi: https://doi.org/10.2118/206249-MS (**Intl. Journal of Green House Gas Control**, in review).

Geologic CO2 sequestration and CO2 enhanced oil recovery (EOR) have received significant attention from the scientific community as a response to climate change from greenhouse gases. Safe and efficient management of a CO2 injection site requires spatio-temporal tracking of the CO2 plume in the reservoir during geologic sequestration. The goal of this paper is to develop robust modeling and monitoring technologies for imaging and visualization of the CO2 plume using routine pressure/temperature measurements.

The streamline-based technology has proven to be effective and efficient for reconciling geologic models to various types of reservoir dynamic response. In this paper, we first extend the streamline-based data integration approach to incorporate distributed temperature sensor (DTS) data using the concept of thermal tracer travel time. Then, a hierarchical workflow composed of evolutionary and streamline methods is employed to jointly history match the DTS and pressure data. Finally, CO2 saturation and streamline maps are used to visualize the CO2 plume movement during the sequestration process.

The power and utility of our approach are demonstrated using both synthetic and field applications. We first validate the streamline-based DTS data inversion using a synthetic example. Next, the hierarchical workflow is applied to a carbon sequestration project in a carbonate reef reservoir within the Northern Niagaran Pinnacle Reef Trend in Michigan, USA. The monitoring data set consists of distributed temperature sensing (DTS) data acquired at the injection well and a monitoring well, flowing bottom-hole pressure data at the injection well, and time-lapse pressure measurements at several locations along the monitoring well. The history matching results indicate that the CO2 movement is mostly restricted to the intended zones of injection which is consistent with an independent warmback analysis of the temperature data.

The novelty of this work is the streamline-based history matching method for the DTS data and its field application to the Department of Engergy regional carbon sequestration project in Michigan.

Paper 4

Deepthi Sen, Hongquan Chen, Akhil Datta-Gupta, *Inter-well connectivity detection in CO2 WAG projects using statistical recurrent unit models*, **Fuel**, Volume 311, 2022, 122600, ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2021.122600.

Routine well-wise injection and production measurements contain significant information on subsurface structure and properties. Data-driven technology that interprets surface data into subsurface structure or properties can assist operators in making informed decisions by providing a better understanding of field assets. Our machine-learning framework is built on the statistical recurrent unit (SRU) model and interprets well-based injection/production data into inter-well connectivity without relying on a geologic model. We test it on synthetic and field-scale CO_2 EOR projects utilizing the water-alternating-gas (WAG) process.

SRU is a special type of recurrent neural network (RNN) that allows for better characterization of temporal trends, by learning various statistics of the input at different time scales. In our application, the complete states (injection rate, pressure and cumulative injection) at injectors and pressure states at producers are fed to SRU as the input and the phase rates at producers are treated as the output. Once the SRU is trained and validated, it is then used to assess the connectivity of each injector to any producer using permutation variable importance method, wherein inputs corresponding to an injector are shuffled and the increase in prediction error at a given producer is recorded as the importance (connectivity metric) of the injector to the producer. This method is tested in both synthetic and field-scale cases.

The validation of the proposed data-driven inter-well connectivity assessment is performed using synthetic data from simulation models where inter-well connectivity can be easily measured using the streamline-based flux allocation. The SRU model is shown to offer excellent prediction performance on the synthetic case. Despite significant measurement noise and frequent well shut-ins imposed in the field-scale case, the SRU model offers good prediction accuracy, the overall relative error of the phase production rates at most producers ranges from 10% to 30%. It is shown that the dominant connections identified by the data-driven method and streamline method are in close agreement. This significantly improves confidence in our data-driven procedure.

The novelty of this work is that it is purely data-driven method and can directly interpret routine surface measurements to intuitive subsurface knowledge. Furthermore, the streamline-based validation procedure provides physics-based backing to the results obtained from data analytics. The study results in a reliable and efficient data analytics framework that is well-suited for large field applications.

Appendix B. Technical Contribution of the Subcontractor: Battelle Memorial Institute

Subrecipient Agreement No. M1803977 / DOE-NETL Agreement No. DE-FE0031625

Robust Carbon Dioxide Plume Imaging Using Joint Tomographic Inversion of Seismic Onset Time and Distributed Pressure and Temperature Measurements

Battelle Contributions

Prepared by:
Battelle Memorial Institute
505 King Avenue
Columbus, Ohio 43201-2696

Technical POC:
Dr. Srikanta Mishra, Co-Principal Investigator
Email: mishras@battelle.org
(614) 424-5712

Submitted to:
Prof. Akhil Datta-Gupta, Principal Investigator
Email: datta-gupta@tamu.edu

December 30, 2021

Chapter 1.

Overview of Report and Main Findings

Srikanta Mishra

Chapter 1. Table of Contents

1.1 Background

The project ‘Robust Carbon Dioxide Plume Imaging Using Joint Tomographic Inversion of Seismic Onset Time and Distributed Pressure and Temperature Measurements’ was selected by U.S. DOE-NETL for award under FOA 0001725. As per the FOA, key research objectives are “*development of modeling and monitoring methods, tools, technologies that improve the certainty about the position of the CO_2 plume over time.*” This is fully addressed by the objective of our project, which is to ***develop and demonstrate computational- and cost-efficient protocols to image the subsurface CO_2 plume via joint inversion of seismic data and distributed temperature and pressure measurements***.

Our motivation stems from the challenges associated with conventional approaches to CO_2 plume mapping using geophysical monitoring and modeling-based workflows. Because of the indirect relationship between the geophysical attributes and transport properties and also the associated uncertainties in the rock physics model, the geophysical techniques are best suited for relatively low resolution and mostly qualitative estimates of CO_2 saturation distribution in the subsurface. Similarly, current modeling-based plume delineation approaches also suffer from non-uniqueness and data sufficiency issues. Therefore, the focus of this project is to advance the theoretical foundations of the joint tomographic inversion-based CO_2 front tracking protocol (as an alternative strategy for CO_2 plume monitoring) and demonstrate its field-scale applicability.

Texas A&M University, the lead organization in the project, was primarily responsible for the development of tomographic approaches for CO_2 plume mapping in conjunction with distributed pressure, temperature and seismic onset time data. Battelle was primarily responsible for the development of analytical and empirical methods for analyzing transient injection rate and pressure data from point/line sources such as injection and monitoring wells. An additional area of emphasis for Battelle was the use of machine learning for such tasks as inferring reservoir connectivity information from injection-production data, and identifying variable importance for machine learning-based proxy models developed from full-physics simulations. The two organizations also collaborated on the application of the tomographic inversion methodology for a field data set.

This report is a compilation of the technical findings from Battelle’s activities noted above. These are organized in eight different chapters as follows:

1.2 Chapter 2 – Detection of CO_2 front location using transient pressure and rate data

Tracking of CO_2 fronts following CO_2 injection is a complicated problem. Normally, it involves the use of time-lapse seismic data, which may or may not be available for a given CO_2 injection project. So, one potential solution is to elucidate the movement of the CO_2 saturation front from the transient pressure response. This approach is based on the fact that: (a) pressure and saturation fronts move in tandem, (b) in-situ saturation data are generally not available, and (c) pressure data are more commonly available at more than one location. The approach presented here is developed based on insights from synthetic (simulated) data using a full-physics model. It involves converting pressure v/s time history at an observation well that has experienced CO_2 breakthrough to an equivalent pressure v/s radial distance profile using the concept of self-similarity. A discontinuity in the pressure response is shown to correspond to the one in the saturation profile at the CO_2-brine interface, which produces an estimate of the CO_2 front location at different times.

The significance of this study is a practical method for front detection from the pressure response alone.

1.3 Chapter 3 – Interpretation of bottomhole temperature data from CO_2 injection projects

Although bottom-hole temperature (BHT) data is routinely collected with bottom-hole pressure (BHP) data during CO_2 injection in geological storage projects, it is seldom analyzed in a quantitative manner. However, the BHP and BHT signals do exhibit correlated behavior. This study proposes a practical approach for using BHT data from warmback (no injection) periods, in conjunction with the amplitude of BHP and BHT changes during the preceding injection event, to estimate formation permeability. The methodology is demonstrated using data from the AEP Mountaineer project in USA where CO_2 injection was carried out into a saline aquifer. Estimated permeability values from the proposed approach are in good agreement with those obtained from transient pressure analysis of BHP data from the falloff period.

The significance of this study is an analytical method for independent estimation of permeability (from temperature data) that can be used as input for detailed inversion studies.

1.4 Chapter 4 – Injectivity index – a powerful tool for characterizing CO_2 storage reservoirs

Injectivity index is a powerful tool that can be easily calculated from field injection rate and pressure data. Theoretically, injectivity index is linked to reservoir properties such as permeability, reservoir thickness, and size of the reservoir. Consequently, injectivity index data can serve as a powerful tool to assess changes in reservoir quality and performance. In this article, we have looked at field data from notable global carbon sequestration projects in order to identify a correlation between injectivity index (J) and the product of reservoir permeability and thickness (kh). Results suggested that there is a linear relationship between J and kh with field data confirming to lower and upper bounds of J = 0.03*kh and J = 0.23*kh respectively. The average trend was calculated to be J = 0.08*kh. This correlative approach can be extremely beneficial when there is a need to estimate pressure buildup corresponding to a target injection or vice-versa. The approach can also be used to understand reservoir flowability and performance at a broader scale prior to the need for expensive or time-consuming flow and pressure transient testing.

The significance of this study is a practical tool for empirical analysis of field data leading to an independent estimate of permeability which can be used an as input for detailed inversion studies.

1.5 Chapter 5 – A screening model for predicting injection well pressure buildup and plume extent in CO_2 geologic storage projects

A new screening model is presented for predicting injection well pressure buildup and CO_2 plume migration. The model requires only limited information and is quite accurate (compared to detailed simulation results). The screening model consists of two correlations: one for injectivity index in terms of the slope of the CO_2 fractional flow curve, and the second for total storage efficiency within the footprint of plume as a function of gravity number and slope of the CO_2 fractional flow curve. Using these two correlations and a knowledge of some basic reservoir characteristics and estimates of fluid properties from correlations, the injection-well pressure buildup and CO_2 plume extent in the formation can be readily estimated. The new correlations show a good match with the results of the underlying simulations, and also provide good agreement with independent calculations for two example problems.

The significance of this study is a practical tool for quick-look analysis and design of what to expect from field observations following CO_2 injection operations. Such models can also assist project developers during project planning, and help regulators perform simple checks against detailed numerical models.

1.6 Chapter 6 – Effect of injection well inclination on injectivity and pressure response

This study examines the impact of injection well inclination (i.e., vertical v/s horizontal well) on well injectivity and monitoring well pressure response. Simulations are carried out to investigate the injectivity index corresponding to horizontal wells of different lengths and different values of reservoir permeability. A correlation is developed for the enhancement in injectivity (compared to the reference vertical well case) as a function of lateral (horizontal well) length and permeability. The difference in pressure response corresponding to a bottomhole pressure gauge located beyond the edge of the lateral for horizontal and vertical well scenarios is examined.

The significance of this study is a practical correlation for assessment of enhanced injectivity index in horizontal wells that can be used for preliminary design purposes.

1.7 Chapter 7 – Field data for imaging pressure and temperature in the subsurface using tomographic methods

As part of this broader research project, the Texas A&M (TAMU) team is developing robust modeling and monitoring technologies for imaging and visualization of the CO_2 plume using routine pressure/temperature measurements. In their approach, they first extend the streamline-based data integration approach to incorporate distributed temperature sensor (DTS) data using the concept of thermal tracer travel time. Then, a hierarchical workflow composed of evolutionary and streamline methods is employed to jointly history match the DTS and pressure data. Finally, CO_2 saturation and streamline maps are used to visualize the CO_2 plume movement during the sequestration process.

This chapter describes the data that was assembled by Battelle to support the above-mentioned objectives, viz: (a) Information about the geologic setting, (b) Oil production and CO2 injection history, (c) Bottomhole pressure and temperature data, (d) Distributed temperature sensing (DTS) data, (e) Numerical model developed by Battelle to integrate geologic, production and injection data, and (f) Additional data types used for model validation (blind testing).

The significance of this activity is a complete field dataset for validation of the TAMU methodology.

1.8 Chapter 8 – Machine-learning based analysis of time-dependent injection-production data

This chapter investigates several data-driven models for charactering reservoir connectivity and forecasting injection-production response from the viewpoints of time series forecasting and regression forecasting. While time series forecasting provides a much better prediction accuracy, regression forecasting typically has a better interpretability in the connectivity between injection and production wells. The recommended model for time series forecasting is the ANN not only because ANNs in general provide the best prediction accuracy but because this ANN has a relatively simple structure so it is possible to open its black-box neural networks to understand the connectivity between injection and production wells. For regression forecasting, the recommended model is the regression model with constraints. Even though the regression model without constraints has a slightly better performance in prediction accuracy, the regression model with constraints has a much better interpretability; specifically, its coefficients can be treated as the proportion contributed from each injection well to that production well.

The significance of this study is a workflow for extracting useful insights from data-driven models of injection-production behavior in subsurface systems.

1.9 Chapter 9 – Evaluating variable importance in time-dependent black box “proxy” models: a comparison of strategies

Measuring variable importance for computational models is an important task in many applications. It is always desirable to have a strategy that works for any model and could uncover the key predictors in the modeling. In this study, we first review several commonly used variable importance strategies that are compatible with all machine learning or black box models and provide a comparative assessment of these strategies using an example from a subsurface geoscience application. Furthermore, we present a framework for making comparisons not only within but also between different time points for time-dependent models. We propose the relative importance score (RIS) and uncertainty importance factor (UIF), which allow users to intuitively interpret how variable importance changes over time.

The significance of this study is a comparative assessment of various strategies for determining the importance of variables in a black-box model, in addition to two metrics to facilitate an understanding of the time-dependent nature of the importance scores.

Chapter 2.

Detection of CO_2 Front Location Using Transient Pressure and Rate Data

Srikanta Mishra, Manoj Valluri, Priya Ravi Ganesh

Chapter 2. Table of Contents

Chapter 2. List of Figures

2.1 Background

Tracking of CO_2 fronts following CO_2 injection is a complicated problem. Normally, it involves the use of time-lapse seismic data, which may or may not be available for at a given CO_2 injection project. So, one potential solution is to elucidate the movement of CO_2 saturation front from transient pressure response. This approach is based on the fact that: (a) pressure and saturation fronts move in tandem, (b) in-situ saturation data are generally not available, and (c) pressure data are more commonly available at more than one location. Our approach is to use simulations to generate synthetic data and utilize that for process understanding and methodology development that could lead to an inexpensive plume tracking tool.

We simulate constant-rate CO_2 injection from a vertical well into a saline aquifer using 2-D radial geometry, a logarithmic expanding grid, and a compositional model. Pressure and saturation profiles are extracted at different radial locations at multiple times. Pressure response is also extracted at injection well and two observation points (with/without CO_2 breakthrough). The examination of these data is presented next.

2.2 Saturation Profiles

Figure 2-1 shows the saturation profiles (i.e., values at different radial location) at different times following injection. Note the logarithmic scale of the x-axis (radial direction). The saturation front appears to stretch in time and space in a self-similar manner.

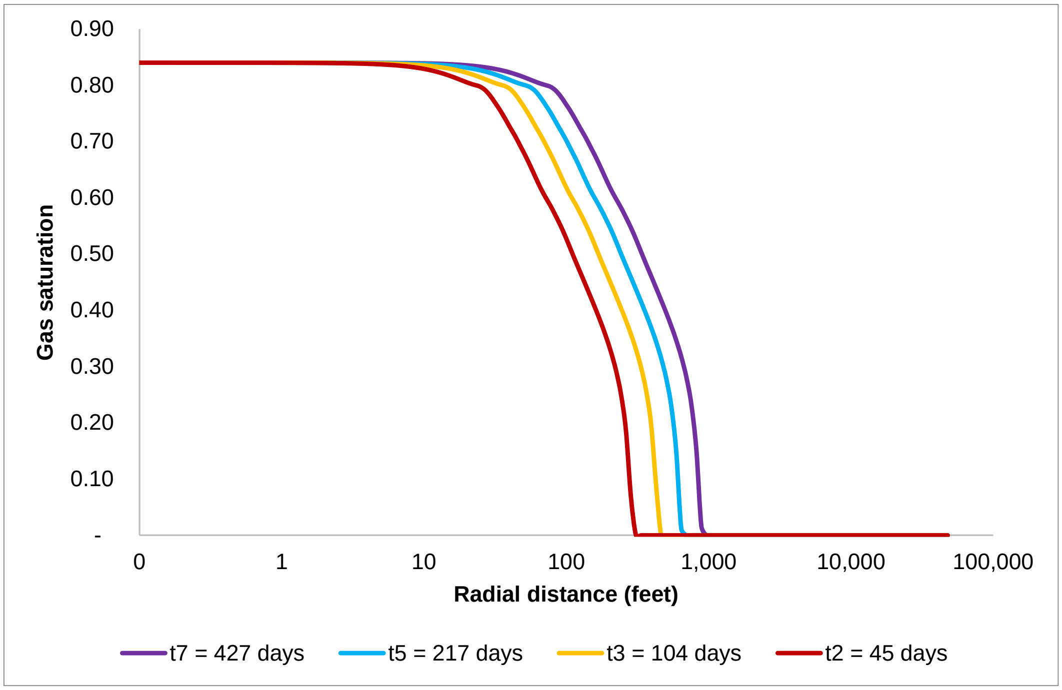


Figure 2-1. Saturation profiles at different times.

As is normally done in advection-dispersion problems, we can use the Boltzman similarity variable r^2/t as a correlating parameter to collapse all curves (as shown in Figure 2). This variable characterizes the propagation velocity of CO_2 plume. Also note the sharp discontinuity at r^2/t = 2050 ft2/day which indicates the location of the CO_2-brine interface.

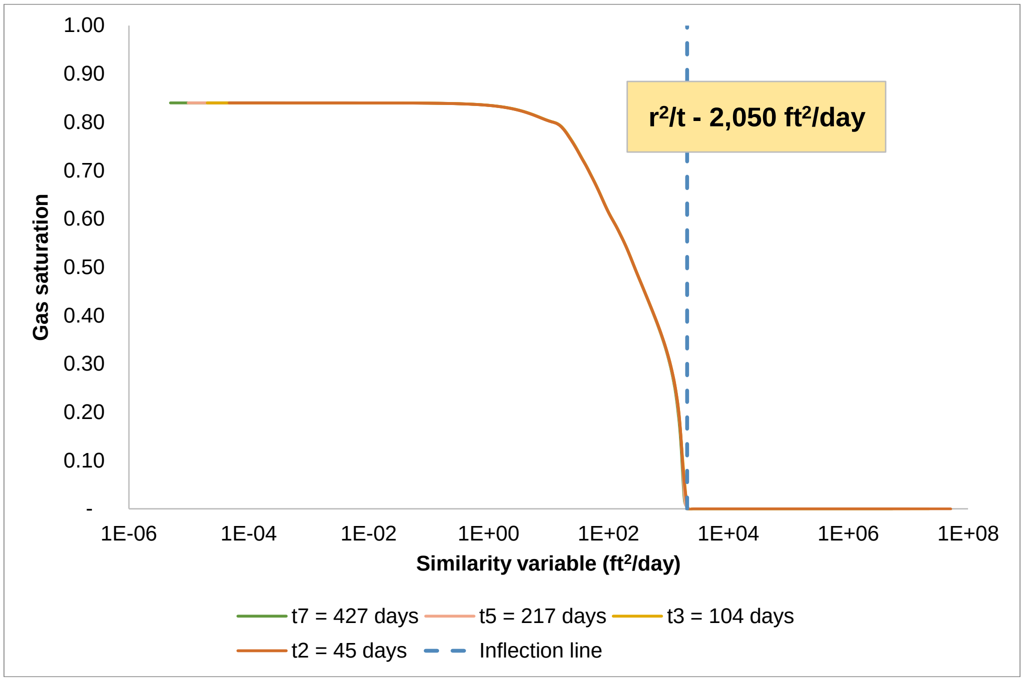


Figure 2-2. Use of Boltzman variable to collapse all saturation profiles into a single curve.

2.3 Pressure profiles

Figure 2-3 shows the pressure profiles (i.e., values at different radial location) at different times following injection. Note the logarithmic scale of the x-axis (radial direction). The pressure front appears to stretch in time and space in a self-similar manner much like the saturation front, albeit significantly further into the formation.

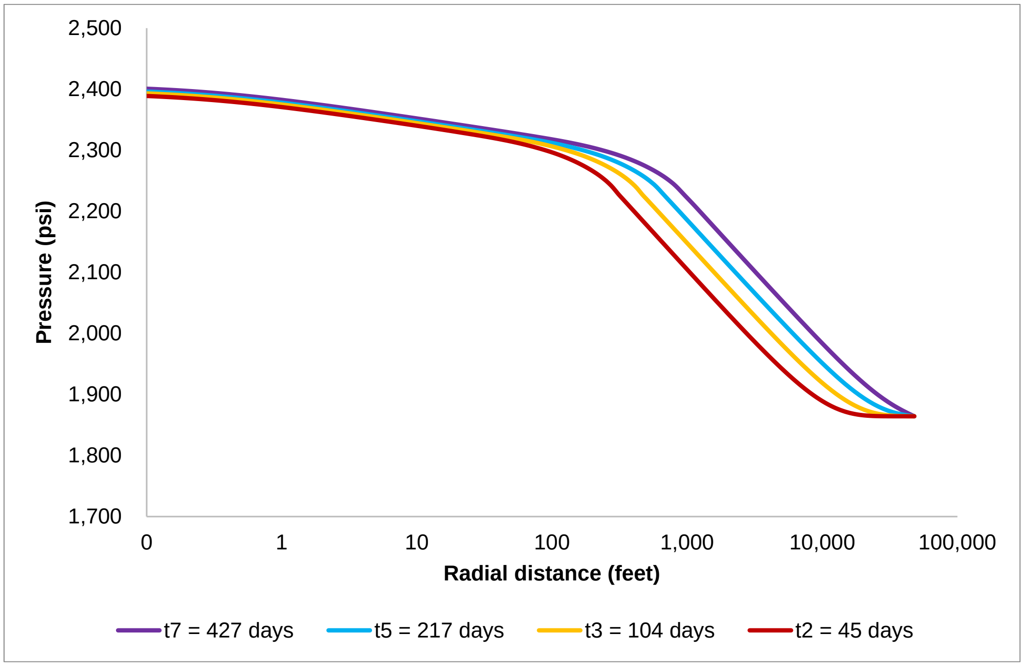


Figure 2-3. Pressure profiles at different times.

Here also, we can use the Boltzman similarity variable r^2/t as a correlating parameter to collapse all curves (as shown in Figure 2.4). For consistency with the saturation profiles, the value of the similarity variable at the location of the saturation discontinuity corresponding to the CO_2-brine interface is shown as a dashed line in this figure.

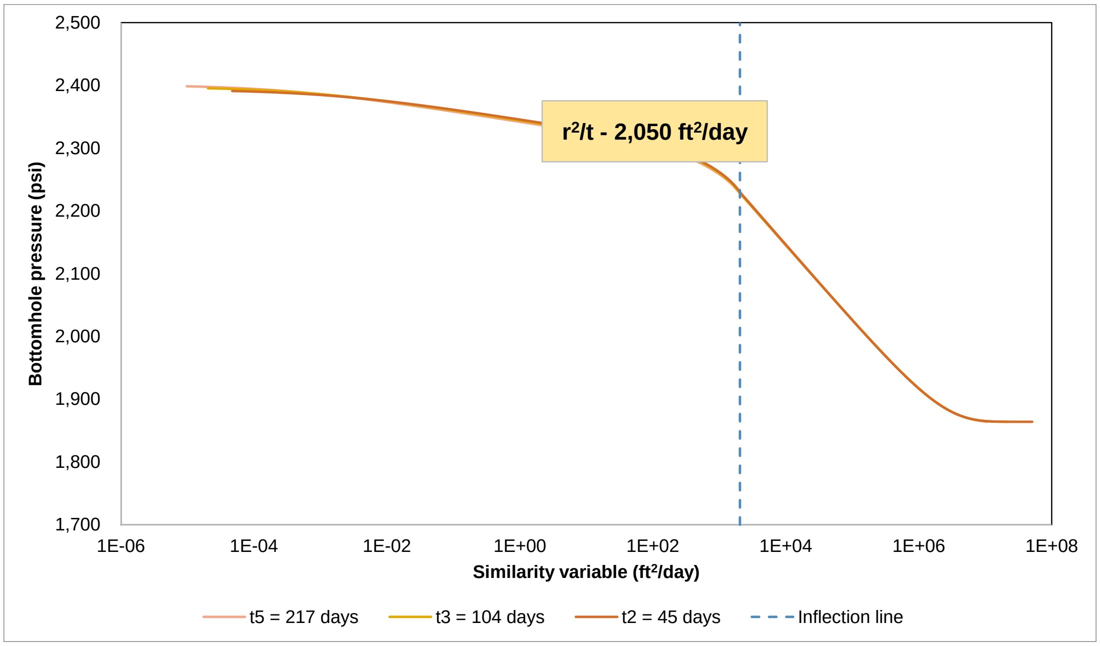


Figure 2.4 Use of Boltzman variable to collapse all pressure profiles into a single curve.

Although it is not obvious that there is a discontinuity in the pressure profile at the same location, such a discontinuity does exist and this will be verified in the next section. This suggests that if the pressure profile were available at different times, one could reduce them using the Boltzman variable, estimate the value of r^2/t at the location of discontinuity and determine the plume boundary from that value.

2.4 Reconstruction of Pressure Profiles from Pressure Time History

Since pressures are typically measured at the injection well and a few monitoring wells, it is not possible to generate the full pressure profile within the system. However, given the demonstrated power of r^2/t as the correlating group in the above-mentioned figures, we examine whether the pressure versus radial location (P ~ r) response can be recovered from the pressure versus time (P ~ t) response. This will be done using data from r = 0.25 ft (injection well) that is always inside the CO_2 plume, r = 137 ft (monitoring well that sees CO_2 breakthrough at some time during the injection history), and r = 5792 ft (monitoring well that is always outside the CO_2 plume).

Figure 2-5 shows the reconstructed pressure profile based on the pressure-time response at r=0.25 ft (in symbols). This is overlain on the full pressure profile calculated at different times and then collapsed into a single curve using r^2/t is the correlating group. The reconstructed values fall on the original pressure profile curve, but the data points are always behind the inflection point corresponding to the location of the CO_2-brine interface.

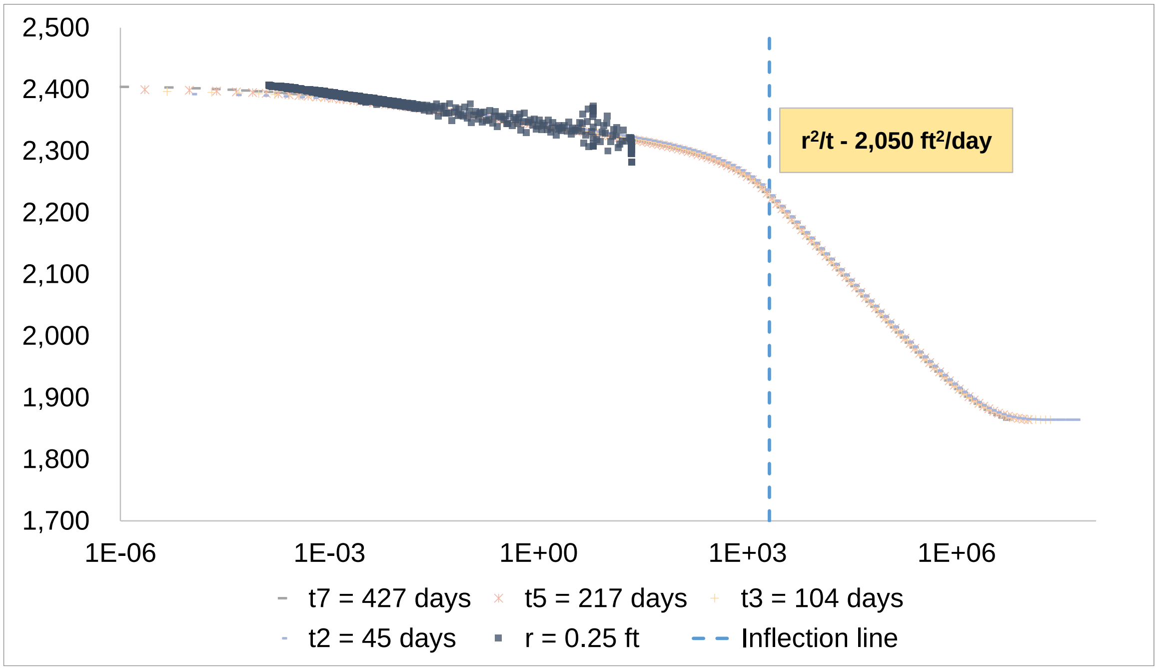


Figure 2-5. Pressure data (y-axis) from different radial locations in light symbols as a function of Boltzman variable r^2/t (x-axis) with overlay of pressure data converted from time domain at r=0.25 ft (dark symbols).

Next, in **Figure 2.6** we show the results for r=137 ft, a location which was initially in the brine-filled region and later became part of the invading two-phase region. Here, the location of the inflection point clearly falls within the spectrum of the reconstructed data.

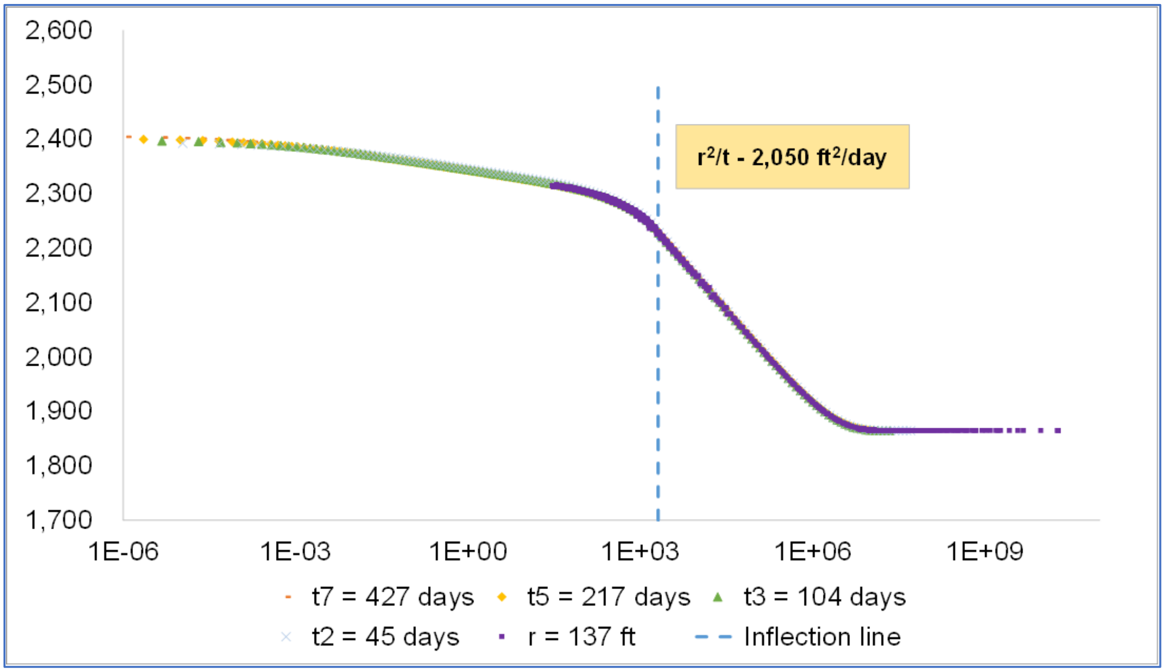


Figure 2.6. Pressure data (y-axis) from different radial locations in light symbols as a function of Boltzman variable r^2/t (x-axis) with overlay of pressure data converted from time domain at r=137 ft (dark symbols).

Finally, we show results from r = 5792 ft, which is always within the brine filled region and is outside the CO_2 plume at all times. Here, the inflection point is outside the range of the reconstructed data.

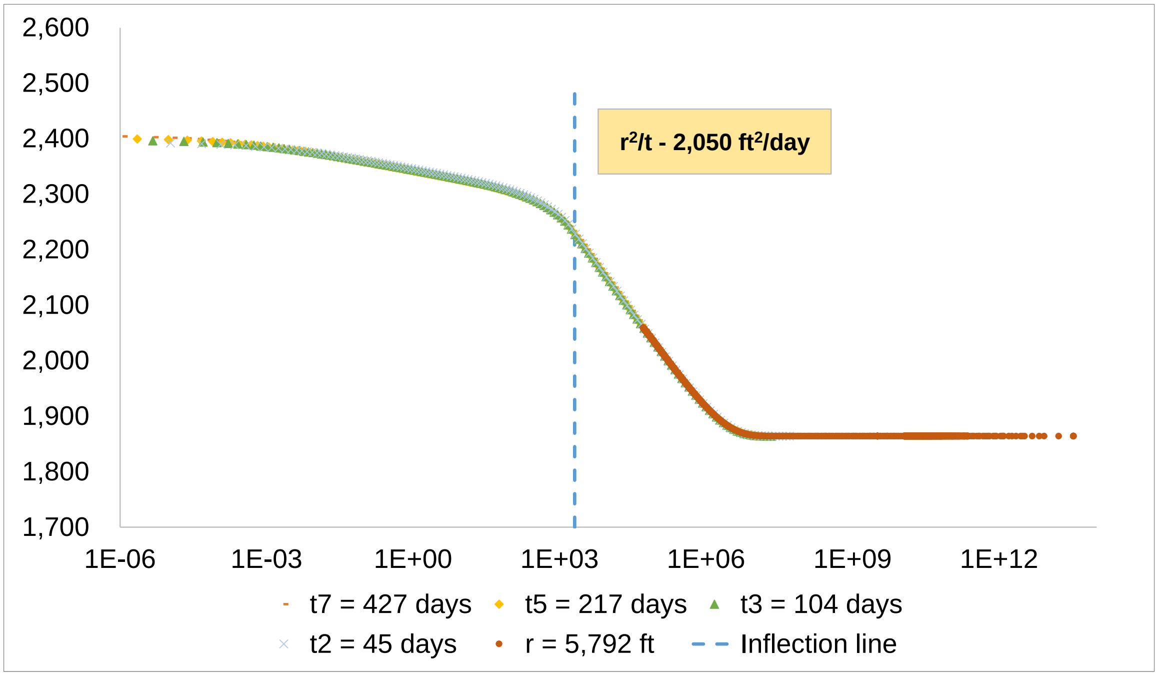


Figure 2-7. Pressure data (y-axis) from different radial locations in light symbols as a function of Boltzman variable r^2/t (x-axis) with overlay of pressure data converted from time domain at r=137 ft (dark symbols).

2.5 Locating the Saturation/Pressure Discontinuity

Thus, we have seen that the location of inflection point only falls within response from r=137 ft, while the first location ($r=r_w$) is always within two-phase region, and the third location (r=5792 ft) always within brine-filled region. Therefore, we need to pick monitoring well location that sees CO_2 breakthrough after some time (e.g., Figure 2.6).

In this figure, the location of the saturation discontinuity is known from simulation results. In a field setting how can this be identified? We propose using a derivative plot to identify this discontinuity, as shown below in Figure 2.8. This involves plotting $dP(r^2/t)$ against the similarity variable and looking for a reversal in slope. This yields a r^2/t value of 2000 ft^2/s at the discontinuity, which compares very well with the simulator input of 2050 ft^2/s. The subscript BL in the figure denotes a Buckley-Leverett shock front, following the original model of frontal displacement for two immiscible fluids in porous media.

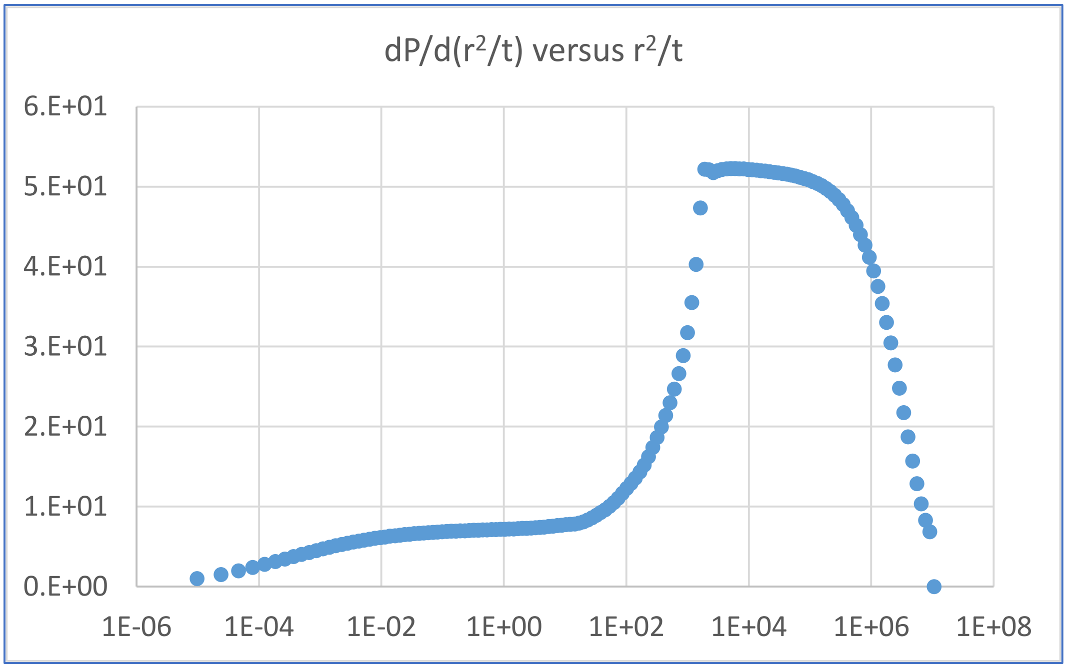


Figure 2-8. Derivative analysis to determine pressure/saturation discontinuity.

In Figure 2-9, the predicted plume radius at different points in time using this value (i.e., $r^2/t|_{front}$ = 2000 ft^2/s) is shown, which agrees very well with the simulator results. We are currently validating this approach with additional datasets.

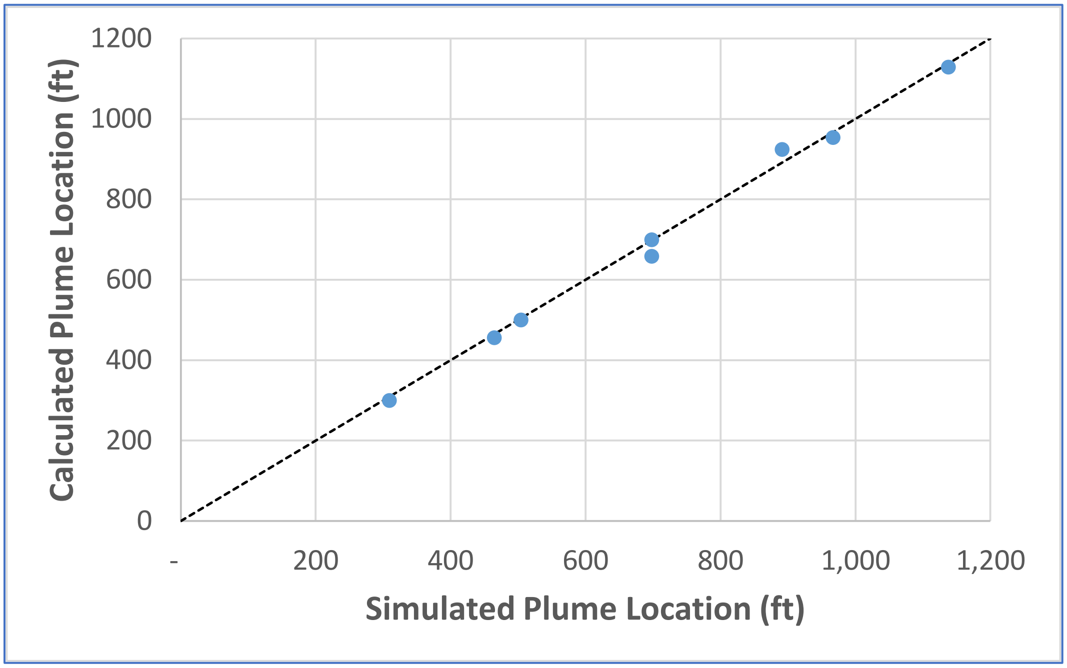


Figure 2-9. Comparison of simulated and predicted plume location based on inferred $r^2/t|_{front}$.

2.6 Summary

We have found that pressure data from a monitoring well location that sees breakthrough at some point after injection can be useful for front tracking process using the following workflow:

- Convert $P(r,t)$ to $P(r^2/t)$
- Pick inflection point from derivative
- This gives r^2/t for BL front
- Can estimate front location at different times

In terms of future work, we propose an extension to variable rate cases (e.g., using deconvolution to generate $P(t)$ for equivalent constant rate response from variable rate $P(t)$ data), and combining this with the material balance time concept for front tracking.

Chapter 3.

Interpretation of Bottom-Hole Temperature Data from CO_2 Injection Projects

Srikanta Mishra, Laura Keister, Sanjay Mawalkar

Chapter 3. Table of Contents

Chapter 3. List of Tables

Chapter 3. List of Figures

3.1 Introduction and Scope

Bottom-hole pressure (BHP) and bottom-hole temperature (BHT) data are routinely collected together as part of monitoring programs in geological storage projects involving CO_2 injection into saline aquifers and depleted oil and gas fields [6]. Since the same gauge can measure both pressure and temperature signals when placed downhole, there is no additional cost in acquiring temperature data. When cold CO_2 is injected into a warm reservoir (as is generally the case), the BHP and BHT signals are generally inverted, i.e., there is thermal cooling as opposed to pressure buildup during injection sequences, and thermal warmback as opposed to pressure falloff during shut-in sequences. As an example, Figure 3-10 shows the correlated nature of the BHP and BHT responses following supercritical CO_2 injection in the AEP Mountaineer project's AEP-1 well [11]. As reservoir properties (e.g., permeability) are known to directly influence the pressure signal, the strong correlation between BHP and BHT signals suggests the potential of inferring these reservoir characteristics from BHT data as well.

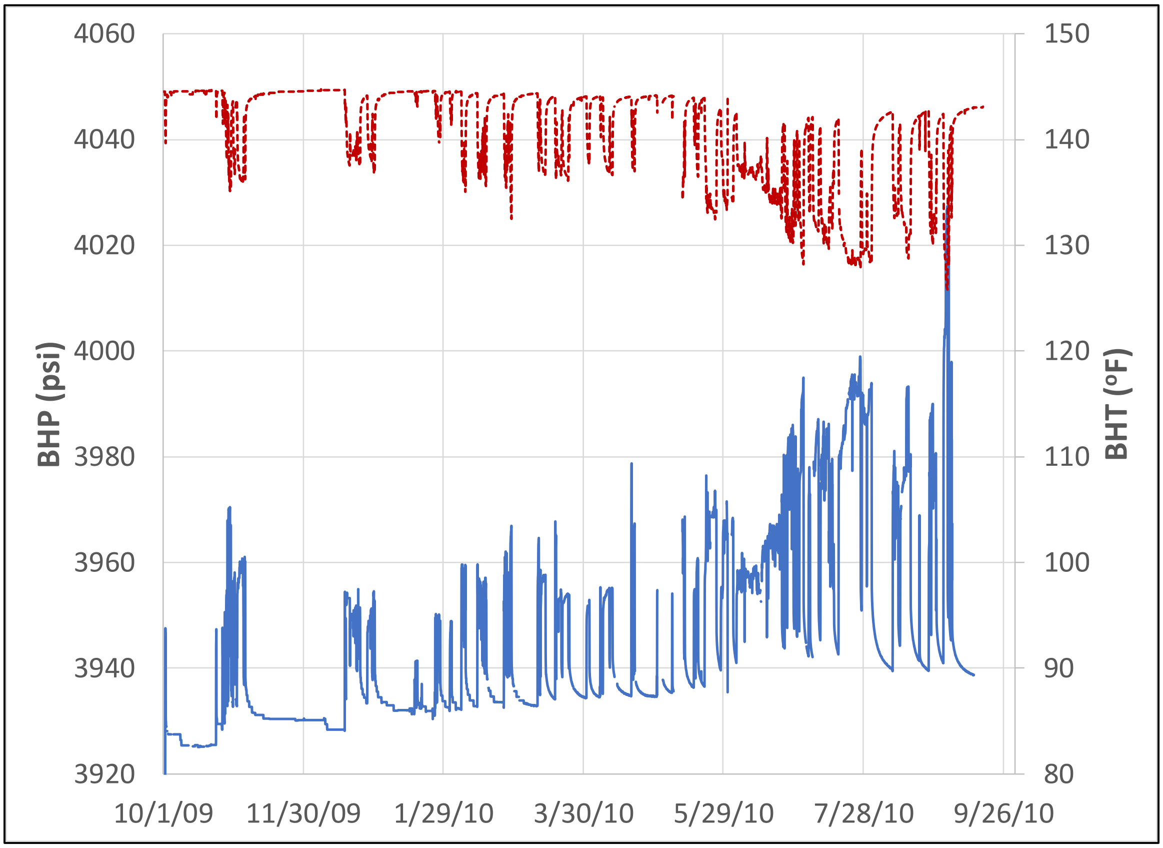


Figure 3-10. Example bottom-hole pressure (BHP) in solid line (left axis) and bottom-hole temperature (BHT) in dashed line (right axis) from AEP Mountaineer project (well AEP-1) showing inverse correlation between BHP and BHT signals (*conversion factors: 1 psi = 6.895 kPa; x ^{o}F = (x-32)*5/9 ^{o}C*).

In CO_2 geologic sequestration projects, pressure data are generally interpreted using analytical models [4, 27, 18]; or used as inputs to numerical reservoir models for history matching to determine key formation characteristics such as permeability [7, 5, 25, 26, 14, 19]). However, to the best of our knowledge, there is very little evidence to indicate that BHT data from individual downhole gauges at a fixed location are actually being used in CO_2 storage projects for any meaningful *quantitative* interpretive purposes. It should be noted a few studies have discussed *qualitative* analysis of fiber-optic based distributed temperature sensing or DTS time series data collected along the length of the borehole. Mawalkar [17] studied the DTS data, in conjunction with pressure and geochemical data, to understand CO_2 movement within the injection zone. Nunez-Lopez [21] used DTS data along with pressure observations to qualitatively monitor CO_2 flow within CO_2 injection zones at the inter-well scale and to detect CO_2 leakage in the[overburden.

Several recent studies in the oil and gas literature have addressed the issue of transient temperature analysis under non-isothermal fluid production conditions for quantifying reservoir properties. Some of the relevant studies are briefly summarized here.

App and Yoshioka [3] used both analytical and numerical models to investigate the interplay between pressure drawdown, Joule-Thomson effect, and permeability in determining the temperature response during fluid production from vertical wells. They concluded that the maximum amplitude of thermal response corresponds to high-permeability reservoirs and is bounded by the Joule-Thomson effect.

Muradov and Davies [20] developed analytical equations for early-time, sandface temperature during fluid production in horizontal wells including a complete treatment of Joule-Thomson and adiabatic expansion effects as well as convection and conduction. Of particular interest are the asymptotic solutions that show the influence of the Joule-Thomson effect in correlating temperature and pressure responses (in a manner similar to that predicted by App and Yoshioka [3]).

Onur and Cinar [22] presented equations for analyzing temperature transient data from constant rate flow tests in infinite-acting single-phase oil reservoirs. The development of these analytical solutions is based on the fact that the effects of temperature changes on pressure transient data can be neglected (but not vice versa) so that the pressure diffusivity and thermal energy balance equations can be decoupled. These analytical solutions account for Joule-Thomson cooling, adiabatic fluid expansion, conduction and convection effects.

Mao and Zeidouni [16] developed a similar analytical model for pressure influenced transient temperature response during the production of a slightly compressible hydrocarbon fluid from a vertical well. Using the Laplace transformation technique, they proposed interpretive equations for analyzing temperature data to infer reservoir properties, such as permeability and damaged-zone radius (if appropriate).

Although these studies (and references cited therein) provide practical approaches for BHT analysis in production wells, their application for injection problems does not appear to be as straightforward. The primary reason for this is the fact that during injection, an additional complicating factor is the surface injection temperature. Together with the injection rate, it affects the temperature at which the supercritical CO_2 reaches the sandface and thus influences the overall temperature change during injection. Note that the injection rate also affects the BHP response via permeability. Thus, the interrelationship between BHT, BHP, injection rate and permeability is the topic of interest for our investigations.

Building on these studies, our motivation here is to demonstrate a practical methodology for analyzing BHT data from CO_2 injection projects and show what information about the reservoir can be reasonably extracted from such data. Specific issues of interest include:

- Does the relationship between the total temperature and pressure changes during injection suggest some information about reservoir properties?
- How can the transient warmback data for the shut-in period after injection be analyzed?
- Is it possible to infer the nature of pressure response (i.e., hence, permeability dependence) from the temperature signal?

The paper is organized as follows. First, we describe the BHP and BHT data from the AEP-1 well in the Mountaineer project, along with steps taken to pre-process the dataset for our analysis. Next, we present a steady-state analysis of the BHP and BHT data across all injection sequences. This is followed by transient temperature analysis using the well-established temperature Horner method, and transient pressure analysis also in terms of the well-known pressure Horner method [9]. This step is necessary to provide the ground truth for our proposed approach, which is presented in the following section. The paper ends with a discussion of the results, its implication for routine analysis of BHT data, and possible improvements in the suggested approach.

3,2 Description of AEP-1 Dataset

The dataset of interest in the present study is from the 20 MW CO_2 capture and storage Product Validation Facility (PVF) project at American Electric Power's Mountaineer Plant in West Virginia, USA [11]. The Mountaineer CO_2 injection system consists of two injection wells—one in the Copper Ridge dolomite formation and one in the Rose Run sandstone formation, and three deep observation wells that were operational between October 2009 and May 2011 [18]. Here, we consider data from the AEP-1 well which is drilled into the Copper Ridge zone consisting of vuggy high porosity and permeability intervals at ~2500 meters depth.

Figure 3-11 shows the location of the wells and the stratigraphic column at the site.

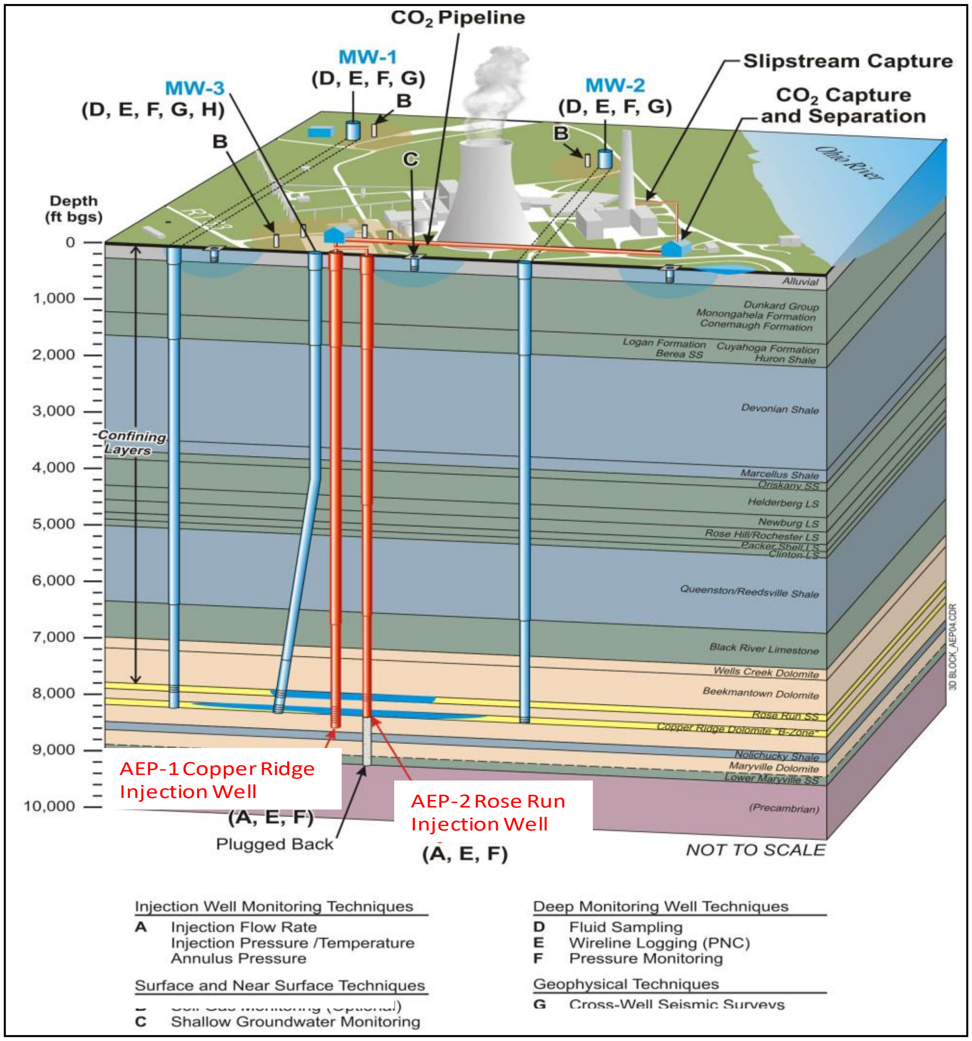


Figure 3-11. Well layout and stratigraphic column at AEP Mountaineer Plant CO$_2$ capture and storage project.

Approximately 27000 metric tons (MT) of CO$_2$ was injected into the Copper Ridge dolomite formation in well AEP-1 with pressure and temperature monitoring undertaken using bottomhole gauges. The injection rates were highly variable because of operational issues with the capture plant. Therefore, as described in Mishra [18], the first step was to simplify the rate history by aggregating all injection (or shut-in) sequences less than 1000 minutes into the previous shut-in (or injection) sequence. For AEP-1, this resulted in 21 injection sequences ranging from 2628 minutes (~1.8 days) to 41406 minutes (~29 days). The average rate for each sequence was calculated to ensure that the cumulative injected volume during each sequence was matched, which yielded the simplified rate history shown in Figure 3-12.

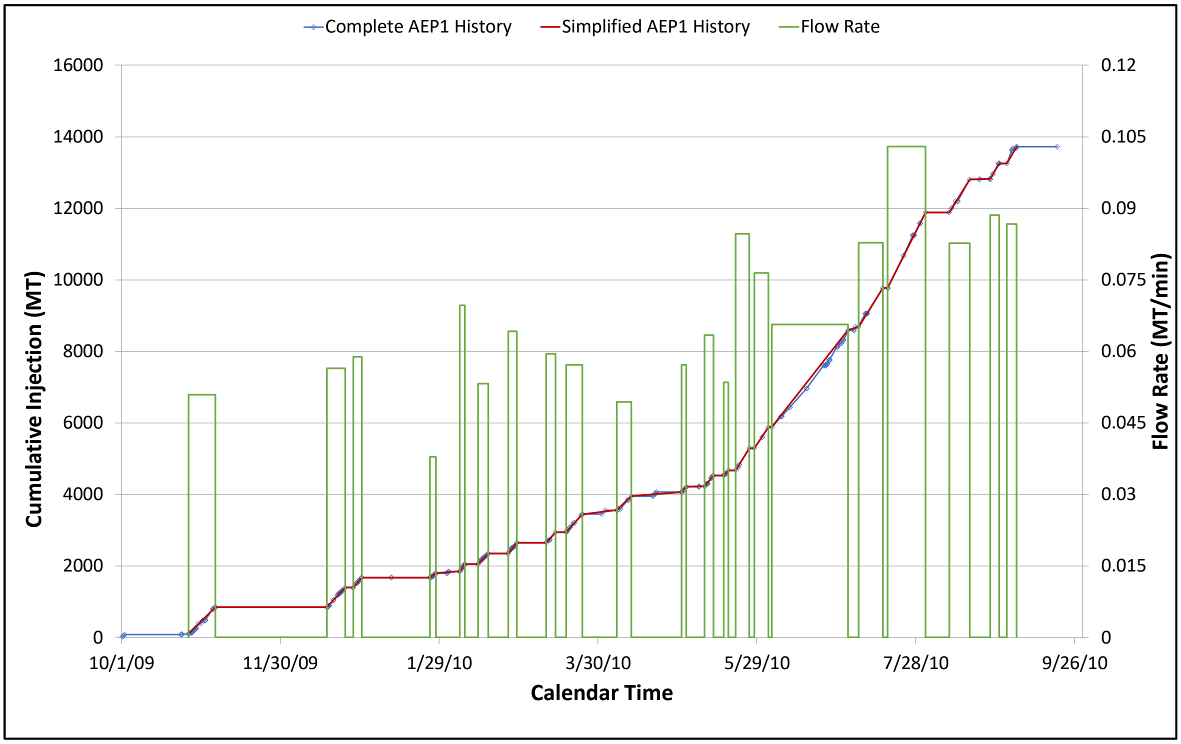
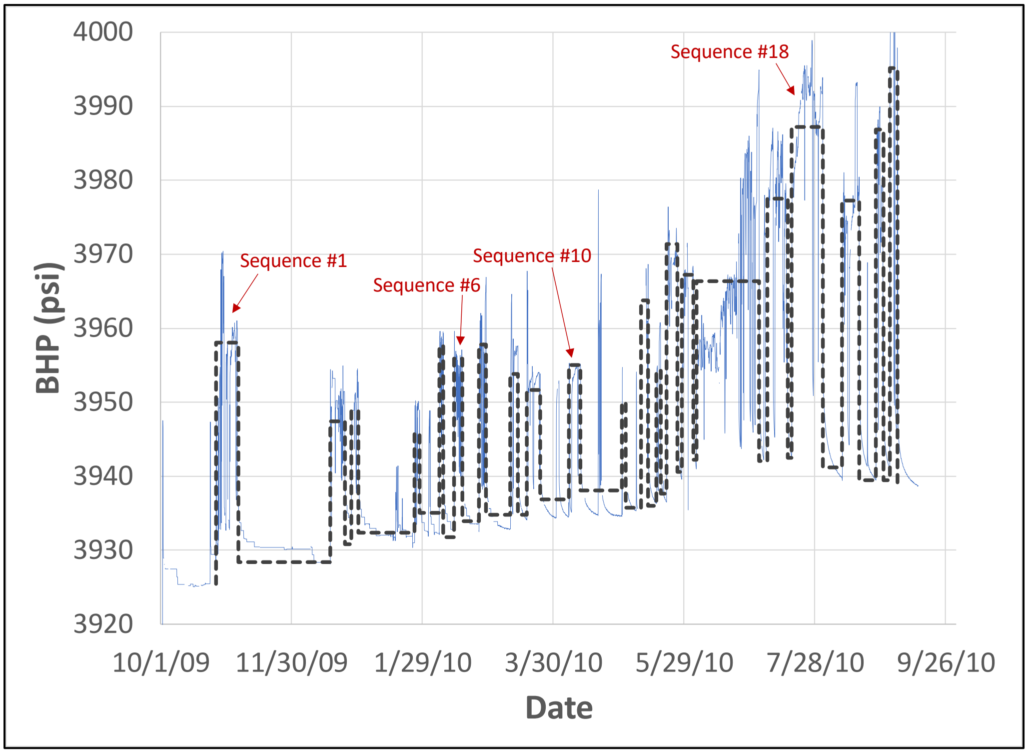


Figure 3-12. Simplification of flow rate history into equivalent constant rate sequences.

While the BHP and BHT data shown earlier in Figure 3-10 are available at a high degree of granularity (i.e., once every minute, re-sampled to once every hour), for the steady-state analysis it is useful to discretize these time series into a series of rectangular pulses reflecting the corresponding quasi steady state (QSS) pressure and temperature changes. Furthermore, the starting and end times associated with these ΔP and ΔT values should be consistent with the simplified rate history developed earlier. Figure 3-13 shows the continuous pressure and temperature history traces juxtaposed with the corresponding rectangular pulse approximations. Henceforth, these quantities will be referred to as ΔP_{QSS} and ΔT_{QSS}.



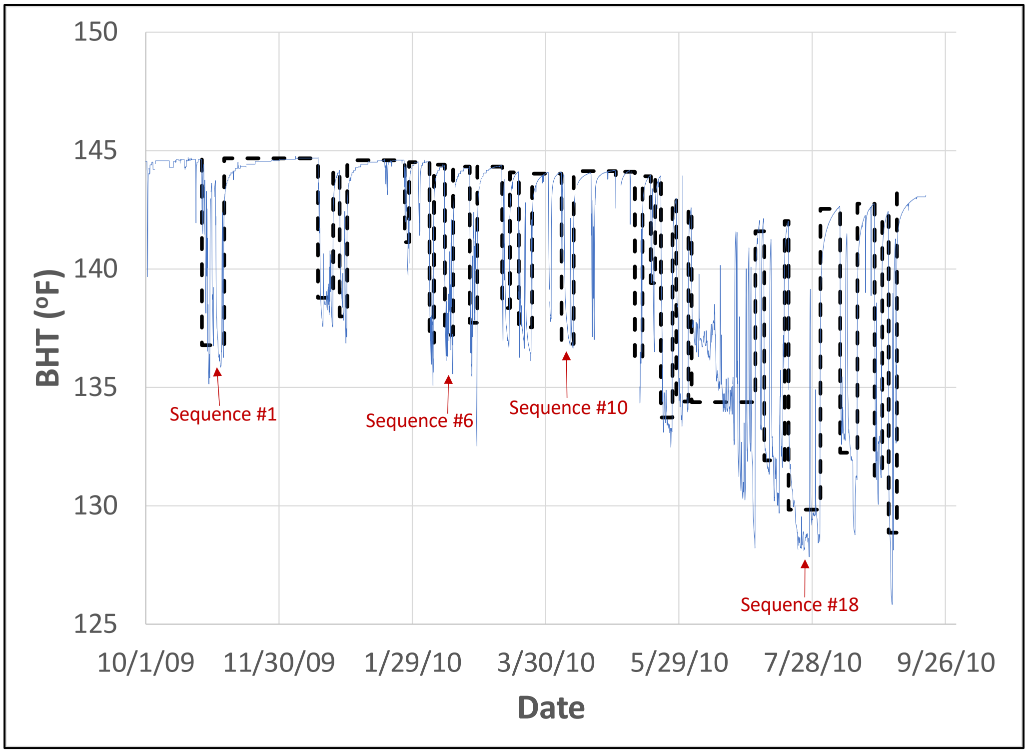


Figure 3-13. Approximation of observed BHP and BHT history (blue line) by rectangular pulses (black dashed line) based on the simplified rate history shown earlier *(conversion factors: 1 psi = 6.895 kPa; x °F = (x-32)*5/9 °C).*

3.3 Steady State Analysis

App and Yoshioka [3] presented a single-well steady-state radial thermal model, based on conservation of energy, conservation of mass and Darcy's law, to describe temperature changes following production of a single-phase fluid. They concluded that the upper bound for the final temperature change is the Joule-Thomson (JT) coefficient. They also found that the temperature change could be attenuated under low-permeability conditions when conductive heat transfer becomes relatively more important compared to convective heat transfer.

Table 3-1 shows the DPQSS and DTQSS values for all 21 sequences and the associated average injection rate. Note the modest pressure and temperature changes observed due to the low rate of injection and high permeability of the formation.

Table 3-1. Observed quasi-steady pressure and temperature changes for various injection events. *(conversion factors: 1 psi = 6.895 kPa; x ^{o}F = (x-32)*5/9 ^{o}C; 1 ^{o}F/psi = 0.08 ^{o}C/kPa)*

Sequence	Injection time (days)	Injection rate	Pressure change	Temp change	dT/dP
	days	MT/day	psi	^{o}F	^{o}F/psi
1	9.75	73.4	29.6	8.0	0.27
2	6.80	81.2	16.7	5.4	0.32
3	3.25	84.8	17.1	6.6	0.39
4	2.27	54.5	10.7	3.4	0.31
5	2.01	100.3	25.9	7.5	0.29
6	3.93	76.6	22.7	7.4	0.33
7	3.28	92.4	23.0	6.6	0.29
8	3.43	85.6	19.0	5.7	0.30
9	6.10	82.3	14.8	6.5	0.44
10	5.46	71.1	21.6	7.6	0.35
11	1.83	82.3	14.2	3.8	0.27
12	3.30	91.3	27.8	7.8	0.28
13	1.84	77.0	16.6	4.5	0.27
14	5.09	121.8	30.8	9.2	0.30
15	5.33	110.0	25.0	8.2	0.33
16	28.75	94.5	24.3	7.2	0.30
17	9.15	119.2	35.1	10.1	0.29
18	14.26	148.3	51.7	13.8	0.27
19	7.71	119.1	37.8	10.5	0.28
20	3.43	127.6	47.4	11.0	0.23
21	3.71	124.8	56.5	14.4	0.26

A cross-plot of these values (Figure 3-14) indicates a strong linear relationship with a slope of 0.28 ^{o}F/psi (0.023 ^{o}C/kPa) and coefficient of determination R^2 = 0.85 (note—this line is constrained to pass through the origin). As per the model of App and Yoshioka [3], this slope should be equal to the JT coefficient, μ_{JT}. However, at a nominal reservoir pressure of 3940 psi (27166 kPa) and reservoir temperature of 140 ^oF (60 ^oC), μ_{JT} for CO_2 is estimated to be 0.8 K/Mpa or 0.01 ^oF/psi (as extrapolated from the data given in http://www.ddbst.com/en/EED/PCP/JTC_C1050.php). Clearly, this is much lower than the value indicated by the data shown in Figure 3-14 — suggesting that other factors beyond simple Joule-Thomson cooling are in play here.

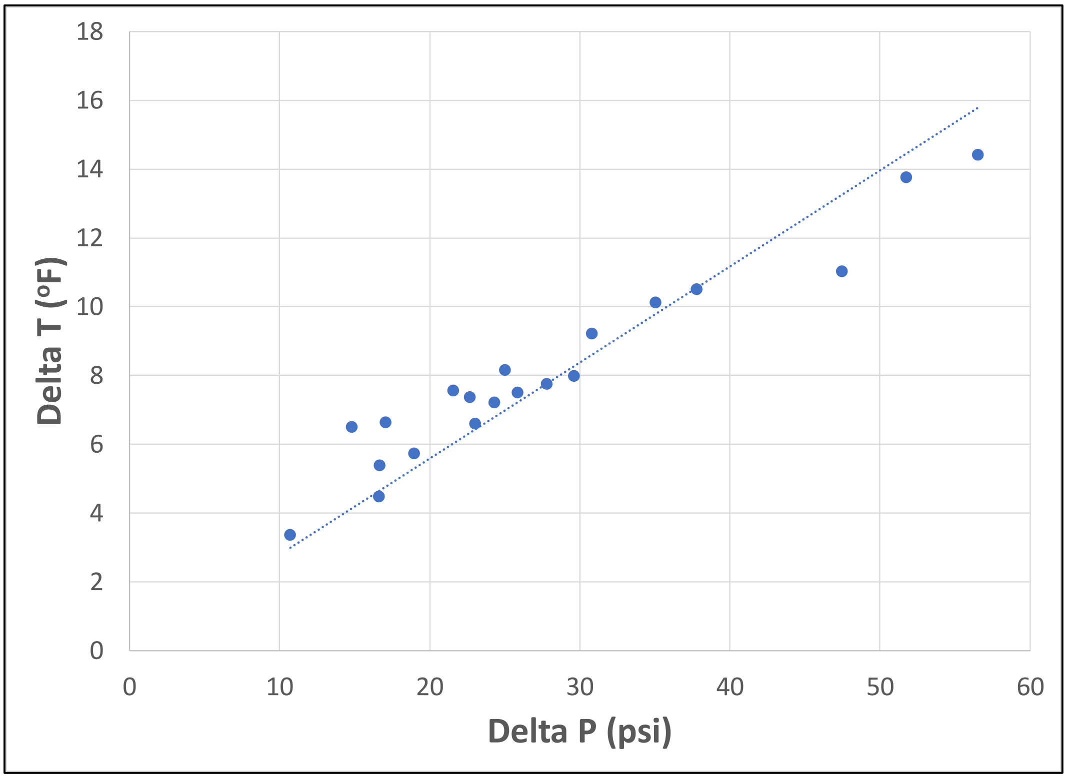


Figure 3-14. Cross-plot of quasi-steady pressure and temperature changes from all injection sequences.
*(conversion factors: 1 psi = 6.895 kPa; x ^oF = (x-32)*5/9 ^oC)*

We hypothesize that during injection, an additional complicating factor is the surface injection temperature. Together with the injection rate, it affects the temperature at which the supercritical CO_2 reaches the sandface and thus influences the overall temperature change (or ΔT_{QSS}) for each sequence. The strong correlation with ΔP_{QSS} suggests that the fluid injection rate is a key linking parameter between the quasi-steady ΔT and ΔP values.

A complete analysis of the injection temperature data would require a numerical model that takes into account all active processes, i.e., Joule-Thomson effect, adiabatic expansion, bulk fluid movement related heat convection, and heat conduction. As noted earlier, some analytical and semi-analytical models are available for such problems under production only conditions [20, 22, 16]. The applicability of these models for injection scenarios is not clear. However, the warmback period following injection is expected to be simpler to analyze because there is no fluid movement and heat conduction should be the primary process that needs to be considered for, as discussed next.

3.4 Transient Temperature Analysis

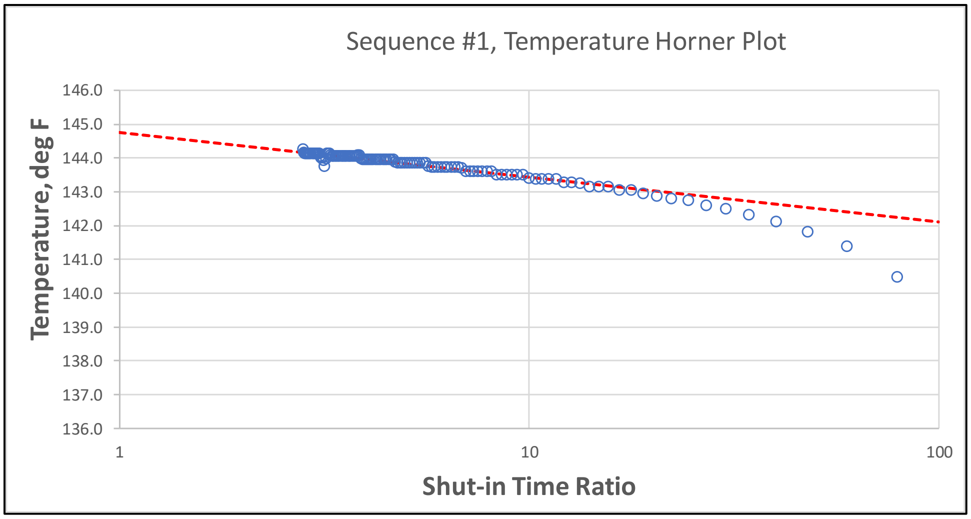
As presented in Dowdle and Cobb [8], Kutasov and Eppelbaum [15], and Gouturbe, *et. al.*, [10], the temperature change for a line source well during shut-in conditions can be expressed as:

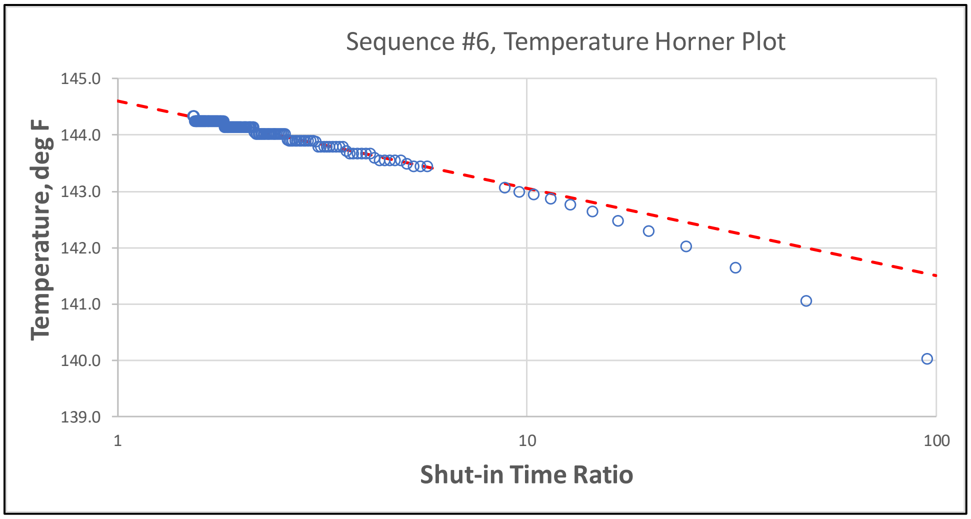
$$\Delta T = T(\Delta t) - T_i = \frac{q_h}{4\pi\lambda_r}\left\{\ln\left(1+\frac{t_{inj}}{\Delta t}\right)\right\} = \frac{q_h}{5.45\lambda_r}\left\{\log\left(1+\frac{t_{inj}}{\Delta t}\right)\right\}$$

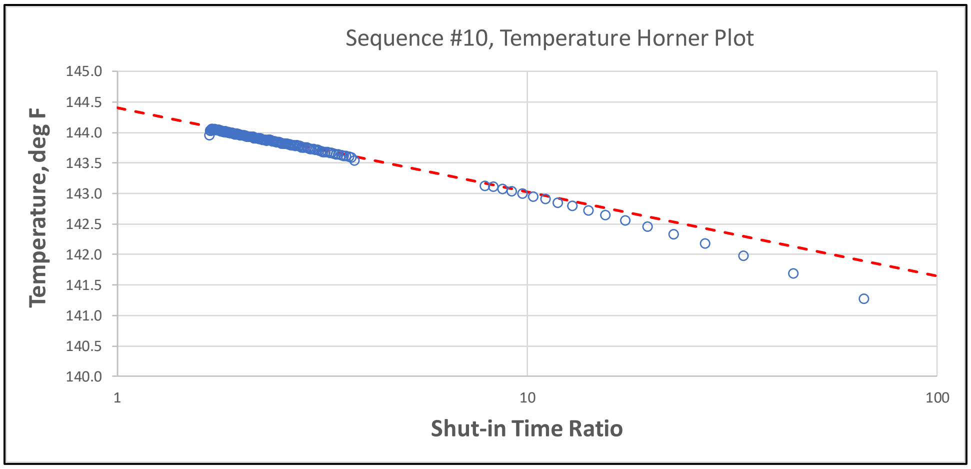
(**Eqn. 3-1**)

where T is time-varying temperature, T_i is initial temperature prior to injection, Δt is shut-in time, t_{inj} is injection time, q_h is heat flux per unit length, and λ_r is thermal conductivity of the in-situ fluid-filled rock (all in consistent units). Eqn. 3-1 is generally referred to as the temperature Horner method based on the similarities between pressure and temperature transient well testing [9], and is the most widely used technique for correcting borehole temperature measurements [10]. Note that Eqn. 3-1 is equally valid for injection as for production events, assuming that the heat flux can be assumed to be constant during injection

For the transient temperature analysis of warmback data in the present study, we pick four sequences, i.e., #1, #6, #10 and #18 (as marked in Figure 3-13), that are separated in time and have varying injection periods. Temperature Horner plots for these sequences are shown in Figure 3-15, where the observed temperature is plotted against the shut-in time ratio, $(1+t_{inj}/\Delta t)$. A straight-line fit to the late-time data suggests that the line source solution as expressed in Eqn. 3-1 is a valid representation of the thermal response. The intercept in each case closely matches the initial temperature (i.e., end temperature corresponding to the DT_{QSS} from Table **3-1**). As evident from Eqn. 1, the slope of the line is proportional to the ratio of unknown heat flux q_h and thermal conductivity λ_r.


Sequence #1, Temperature Horner Plot
Temperature, deg F
146.0
145.0
144.0
143.0
142.0
141.0
140.0
139.0
138.0
137.0
136.0
1
10
100
Shut-in Time Ratio


Sequence #6, Temperature Horner Plot
Temperature, deg F
145.0
144.0
143.0
142.0
141.0
140.0
139.0
1
10
100
Shut-in Time Ratio


Sequence #10, Temperature Horner Plot
Temperature, deg F
145.0
144.5
144.0
143.5
143.0
142.5
142.0
141.5
141.0
140.5
140.0
1
10
100
Shut-in Time Ratio

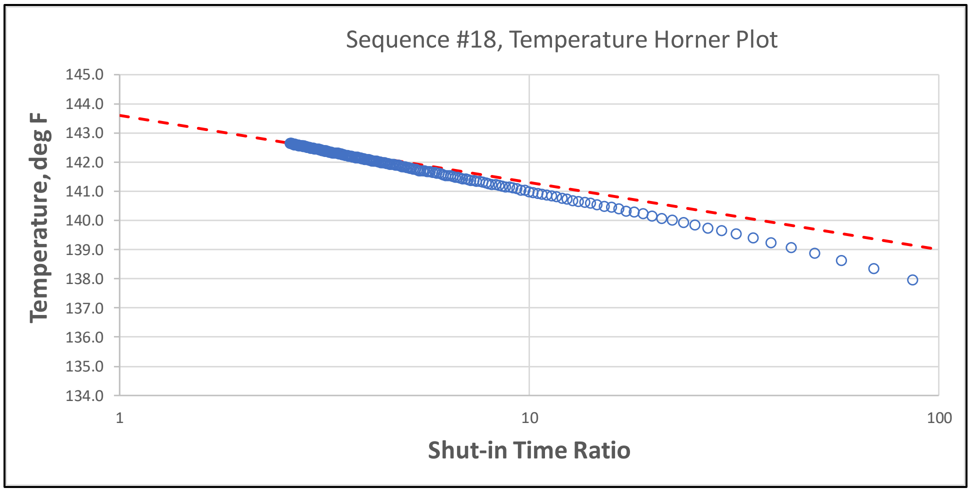


Figure 3-15. Temperature Horner plots for selected warmback sequences. *(conversion factor: x $^oF = (x-32)*5/9\ ^oC$)*

Assuming a rock thermal conductivity value of 2 W/m-K (1.156 BTU/ft-^{0}F) as a representative value for the vuggy dolomite Copper Ridge formation [24], the heat flux q_h can be calculated for each sequence from the slope of the Horner plots using Eqn. 3-1. Table 3-2 shows the calculated q_h values, as well as the observed fluid flux q_f (i.e., injection rate per unit length or effective formation thickness). Here, the surface fluid injection rate has been converted into a downhole flux by assuming a CO_2 formation volume factor (B_{CO2}) of 2.285 ft^3/MCF (0.0023 m^3/sm^3) (Jarrell, *et. al.*, [13]) and an effective formation thickness of 31 ft (10 m) (Mishra, *et. al.*, [19]). Note the strong correlation between the two variables shown in Table **3-2**.

Table 3-2. Correspondence between calculated heat flux and observed fluid flux*(conversion factors: 1 BTU/ft-hr = 0.96 W; 1 ft^3/ft-hr=2.6E-5 m^3/m-s)*

Seq	q_h	q_f
	BTU/ft-hr	Ft^3/ft-hr
1	8.35	4.27
6	9.77	4.47
10	8.66	4.14
18	14.49	8.64

We further explore the issue of this correspondence by noting that from fundamental considerations we can express the heat flux (q_h) associated with fluid injection in terms of fluid flux (q_f) as follows:

$$q_h = \alpha q_f (\rho C)_f \qquad \textbf{(Eqn. 3-2)}$$

where $(\rho C)_f$ is the density-weighted specific heat capacity of the injected fluid and α is a proportionality constant with units of temperature (^{o}F or R). From dimensional considerations, it could be argued that $\alpha=\Delta T$ (i.e., difference between formation and injected fluid temperatures). However, using a representative value of $(\rho C)_f$ = 25.3 BTU/(ft^3-F), and the average values of q$_h$ and q$_f$ from Table 3-2, we get α=0.078, which is significantly different from the temperature differences indicated in Table 3-1. Therefore, we treat α as an unknown constant of proportionality that generally represents the efficiency of the heat transfer process. In fact, our workflow does not require any explicit knowledge of this parameter, as will be shown later.

Combining Eqn. 3-1 and Eqn. 3-2, we get

$$\frac{T(\Delta t)-T_i}{q_f}=\frac{\alpha(\rho C)_f}{5.45\lambda_r}\left\{\log\left(1+\frac{t_{inj}}{\Delta t}\right)\right\} \qquad \textbf{(Eqn. 3-3)}$$

which can be considered as the equation for a rate-normalized (RN) temperature Horner analysis. Eqn. 3-3 suggests that data from multiple sequences can be examined and analyzed using the same plot, since the slope is dependent only on thermal properties of the injected fluid and the in-situ fluid filled rock, i.e.,

$$m_{RN,T}=\frac{\alpha(\rho C)_f}{5.45\lambda_r} \qquad \textbf{(Eqn. 3-4)}$$

This leads us to re-analyze the data shown earlier in Figure 3-15 using rate-normalization, which is presented in Figure 3-16**.** The late time portions of the warmback data from all four sequences fall on a single straight line characterized by a slope of 0.31 ^{o}F/(ft^2/hr)/log-cycle (6.6E3 ^{o}C/(m^2/s)/log-cycle).

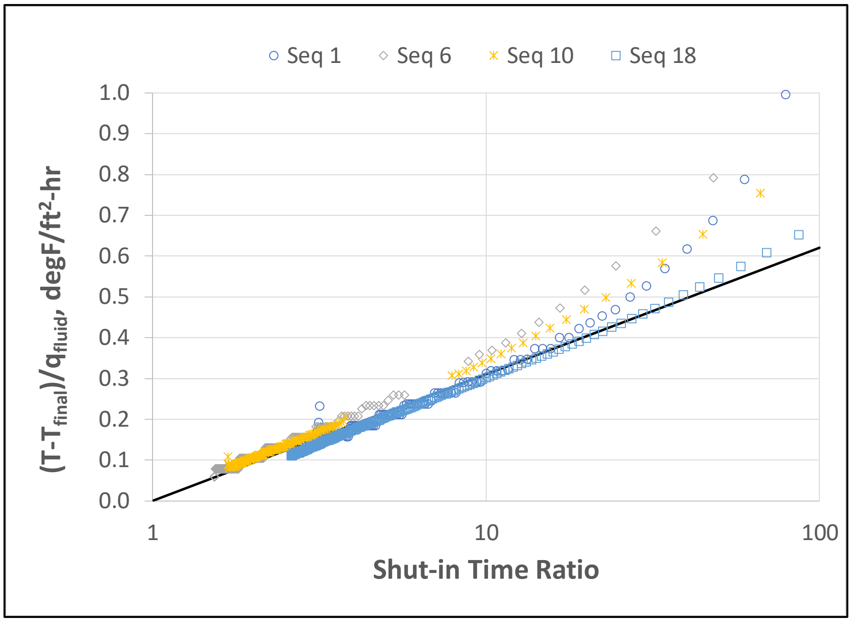


Figure 3-16. Rate-normalized temperature Horner plots for consistency check across selected warmback sequences *(conversion factor: 1 ^{o}F/(ft^2/hr) = 2.1E4 ^{o}C/(m^2/s))*.

The success of rate-normalization in correlating much of the warmback data suggests that the key thermal properties affecting BHT response (i.e., effective heat capacity, which relates fluid flux to heat flux, and thermal conductivity, which determines the rate of heat transfer from the injected fluid to the rock and in-situ fluid) are generally unchanging with time. However, the dependence on reservoir properties such as permeability is missing from this analysis. This will be attempted next by performing transient pressure analyses and seeking a linkage between the pressure and temperature analyses, as discussed next.

3.5 Transient Pressure Analysis

As noted by Mishra, *et. al.*, [18], based on the work on injection well testing summarized by Abbaszadeh [1] and the composite system models of Ramey [23] and Ambastha [2], the late-time pressure fall-off response in the injection well can be expressed as:

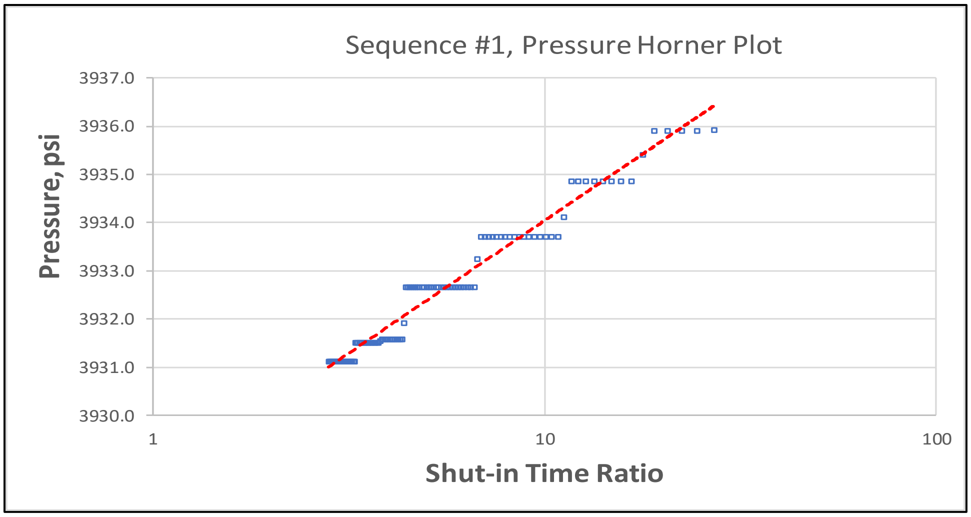
$$\Delta P = P(\Delta t) - P_i \cong \frac{q\mu_w}{4\pi kh}\left\{\ln\left(1+\frac{t_{inj}}{\Delta t}\right)\right\} \qquad \textbf{(Eqn. 3-5)}$$

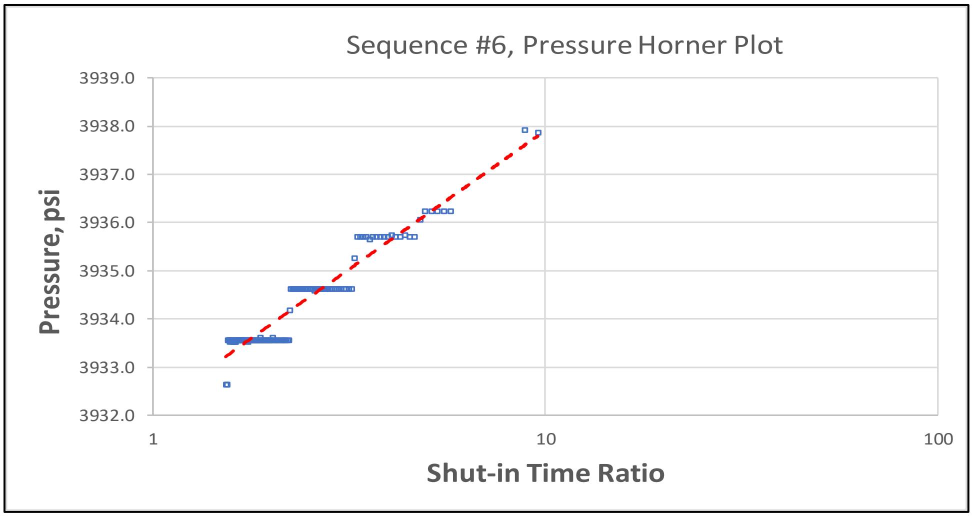
Note that the viscosity of interest here is that of the brine phase, indicating that the pressure response is essentially being influenced by the undisturbed formation. Also, the q in Eqn. 3-5 is in reservoir volumes, which requires converting the mass injection rate at the surface into a sandface rate using the CO_2 formation volume factor, B_{CO2}, at the appropriate pressure and temperature conditions [13]. Using the definition of the fluid flux $q_f = q/h$, we can re-write Eqn. 3-5 as:

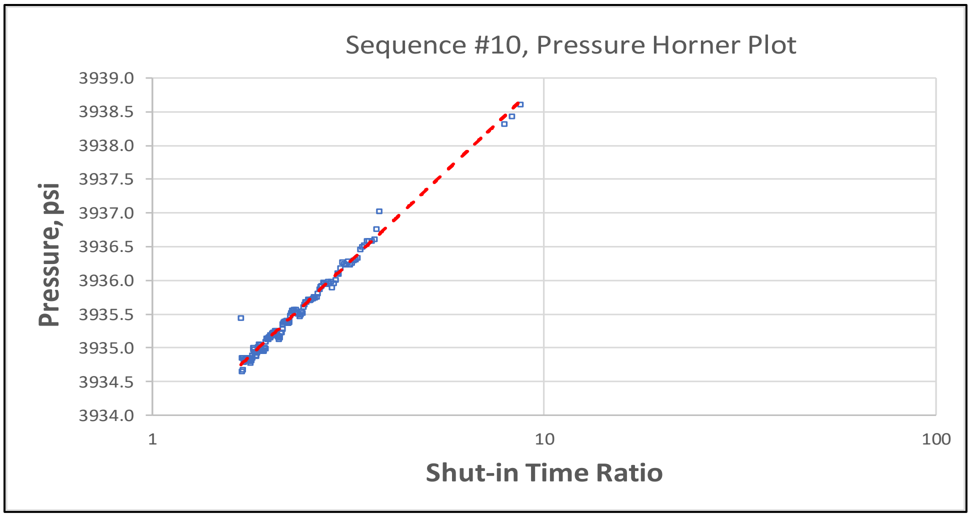
$$P(\Delta t) - P_i \cong \frac{695 q_f \mu_w}{k}\left\{\log\left(1+\frac{t_{inj}}{\Delta t}\right)\right\} \qquad \textbf{(Eqn. 3-6)}$$

where ΔP is in psi, q_f is in $ft^3/ft/hr$, μ_w is in cp, and k is in mD (the constant is 1.26E8 when ΔP is in kPa, q_f is in $m^3/m/s$, μ_w is in cp, and k is in m^2). Eqn. 3-6 thus provides a straightforward way for analyzing the falloff response via a semi-log plot of time-varying shut-in pressure against the shut-in time ratio, $(1+t_{inj}/\Delta t)$.

As in the case of the transient temperature analysis, we pick sequences #1, #6, #10, and #18 for the transient pressure analysis of the falloff data. Pressure Horner plots for these sequences are shown in Figure 3-17. In general, the late-time portions of the data (shown here) indicate strong linear trends—notwithstanding the stair-stepping that possibly reflects measurements with pressure differentials close to the gauge resolution. The straight-line fits to the late-time data suggests that the expected falloff response as expressed in Eqn. 3-6 is a valid representation of the pressure behavior. The intercept in each case closely matches the end pressure used to calculate the ΔP_{QSS} values in Table **3-1**. The slope is proportional to the ratio of fluid flux times brine viscosity and reservoir permeability.


Sequence #1, Pressure Horner Plot
Pressure, psi
3937.0
3936.0
3935.0
3934.0
3933.0
3932.0
3931.0
3930.0
1
10
100
Shut-in Time Ratio


Sequence #6, Pressure Horner Plot
Pressure, psi
3939.0
3938.0
3937.0
3936.0
3935.0
3934.0
3933.0
3932.0
1
10
100
Shut-in Time Ratio


Sequence #10, Pressure Horner Plot
Pressure, psi
3939.0
3938.5
3938.0
3937.5
3937.0
3936.5
3936.0
3935.5
3935.0
3934.5
3934.0
1
10
100
Shut-in Time Ratio

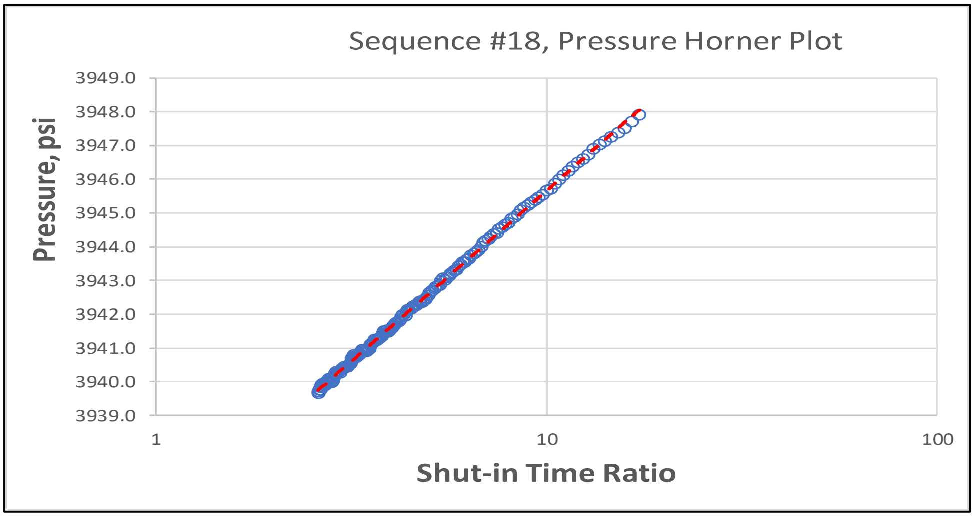


Figure 3-17. Pressure Horner plots for selected falloff sequences *(conversion factor: 1 psi = 6.895 kPa).*

Rearranging Eqn. 3-6 by normalizing with respect to fluid flux yields:

$$\frac{P(\Delta t)-P_i}{q_f} \cong \frac{695\mu_w}{k}\left\{\log\left(1+\frac{t_{inj}}{\Delta t}\right)\right\} \qquad \textbf{(Eqn. 3-7)}$$

which can be considered as the equation for a rate-normalized (RN) pressure Horner analysis. Eqn. 3-7 suggests that pressure data from multiple sequences can be examined and analyzed using the same plot, since the slope is dependent only on intrinsic permeability and brine viscosity, i.e.,

$$m_{RN,P} = \frac{695\mu_w}{k} \qquad \textbf{(Eqn. 3-8)}$$

As before, a rate-normalization Horner plot is constructed per Eqn. 3-7. This is presented in Figure 3-18 with a slope of 1.15 psi/(ft^2-hr)/log-cycle (3e5 kPa/(m^2-s)/log-cycle), which shows a general consistency in late-time pressure response across all sequences. This suggests that the underlying reservoir characteristics, especially formation permeability, are not changing with time.

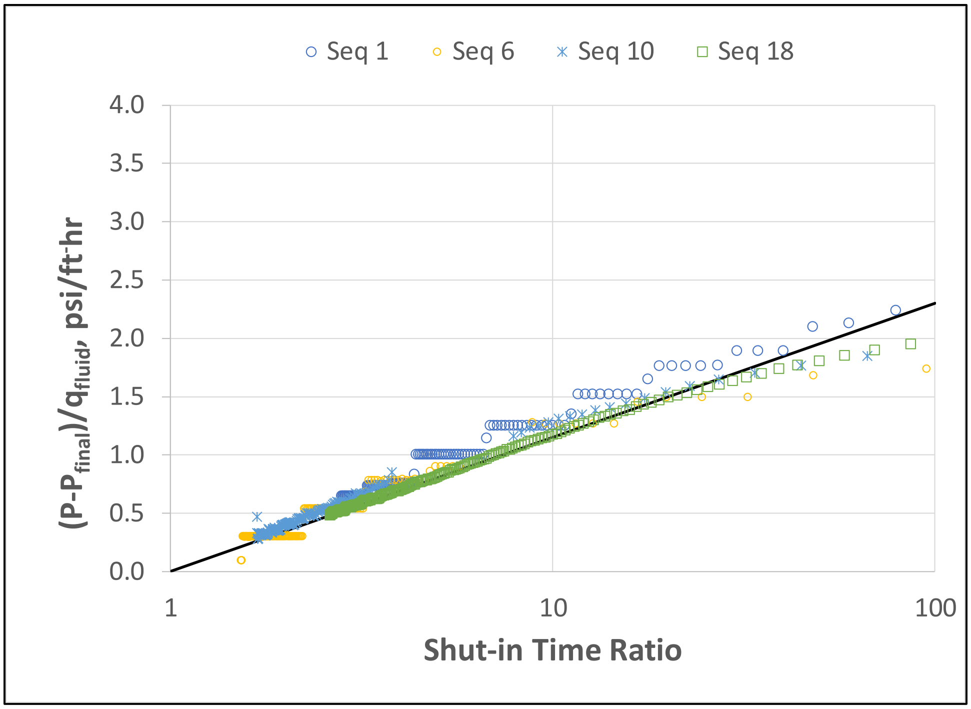


Figure 3-18. Rate-normalized pressure Horner plots for consistency check across selected falloff sequences. (*conversion factor: 1 psi/(ft^2/hr) = 2.7E5 kPa/(m^2/s)*.

Using this slope value and Eqn. 3-8, assuming a brine viscosity of 1.2 cp (calculated at P = 3940 psi (27661 kPa), T=140 ^{0}F (60 ^{0}C), salinity = 300000 ppm) and a formation thickness of 31 ft (10 m), the permeability-thickness product is determined to be kh = 23293 mD-ft (7.0E-6 m^3). This compares very well with the value of 23940 mD-ft (7.2E-6 m^3) obtained by Mishra, *et. al.*, [18] using conventional pressure-derivative based transient pressure analysis methods.

It should be pointed that the purpose of performing a transient pressure analysis of the falloff data here is to provide an independent assessment of the reservoir permeability and determine the extent to which it can be independently determined from the BHT signal (in conjunction with some aspects of the BHP signal). This is discussed next.

3.6 New Approach for BHT Analysis

Our basic hypothesis is that the ratio of quasi-steady state temperature and pressure changes are the same as the ratio from the temperature and pressure Horner plots, as both are essentially reflecting the underlying factors influencing temperature and pressure response. This can be stated as:

$$\frac{\Delta T_{QSS}}{\Delta P_{QSS}} = \frac{m_T}{m_P} = \frac{m_{RN,T}}{m_{RN,P}} \qquad \textbf{(Eqn. 3-9)}$$

where $m_T = m_{RN,T} * q_f$ and $m_P = m_{RN,P} * q_f$. Thus, knowing the ratio of quasi steady state pressure and temperature changes and using the analysis of warmback data in the form of a temperature Horner plot, the slope of the corresponding pressure Horner plot can be estimated via Eqn. 3-9, from which the reservoir permeability can be calculated using Eqn. 3-8. Note that this does not require the direct analysis of the time-dependent pressure fall-off response.

To test this hypothesis for our dataset, we start by noting that the ratio of $\Delta T_{QSS}/\Delta P_{QSS}$ averaged across all sequences is 0.28 ^oF/psi or 0.023 ^oC/kPa (as indicated by Figure 3-14). From the rate-normalized temperature Horner plot (Figure 3-16) we have slope $m_{RN,T}$ = 0.31 $^oF/(ft^2/hr)$/log-cycle or 6.6E3 $^oC/(m^2/s)$/log-cycle, and from the rate-normalized pressure Horner plot (Figure 3-18) we have slope $m_{RN,P}$ = 1.15 psi/(ft^2/hr)/log-cycle or 3E5 kPa/$(m^2$-s)/log-cycle. The ratio of these two values is 0.31/1.15 = 0.27 ^oF/psi (or 6.6E3/3E5 = 0.022 ^oC/kPa), which is in excellent agreement with the value of 0.28 ^oF/psi or 0.023 ^oC/kPa calculated using the quasi steady state rectangular pulse approximation approach. This confirms the hypothesis postulated via Eqn. 3-9.

We verify these ideas further for the four individual sequences of interest as shown below in Table 3-3. The various quantities presented there are as follows:

- q_f = reservoir fluid flux (i.e., injection rate converted to sandface conditions and divided by formation thickness),
- dP $\equiv \Delta P_{QSS}$ = quasi steady state pressure change,
- dT $\equiv \Delta T_{QSS}$ = quasi steady state temperature change,
- slope P $\equiv m_P$ = slope of the pressure Horner plot,
- slope T $\equiv m_T$ = slope of the Temperature Horner plot,
- slope ratio = m_T/m_P,
- dT/dP = ratio of QSS temperature and pressure changes,
- Slope error = difference between slope ratio and dT/dP value,s
- Slope P = slope of pressure Horner plot calculated as per Eqn. 3-9,
- Calculated kh = permeability thickness product from slope P as per Eqn. 3-8.

Table 3-3. Calculation of permeability using new approach.

Seq	t_{inj}	q_{inj}	dP	dT	slope P	slope T	slope	dT/dP	slope	slope P	kh calc
	days	ft^3/ft-hr	psi	deg F	psi/~	F/~	ratio	ratio	error (%)	calc	mD-ft
1	9.75	4.27	29.6	8.0	5.5	1.325	0.24	0.27	10.7	4.91	21764
6	3.92	4.47	22.7	7.4	5.75	1.55	0.27	0.33	17.2	4.76	23465
10	5.46	4.14	21.6	7.6	5.47	1.375	0.25	0.35	28.3	3.92	26432
18	14.25	8.64	51.7	13.8	10.11	2.3	0.23	0.27	14.5	8.65	24997
all	(using rate-normalized plots)				1.15	0.31	0.27	0.28	3.7	1.11	23293

The relevant conversion factors have all been presented earlier. There are two key observations to be gleaned based on the results of Table 3-3. First, the calculated kh values are generally self-consistent with an average of 24166 mD-ft (7.3E-6 m^3), which compares very well with the value of 23293 mD-ft (7.0E-6 m^3) calculated from the rate-normalized pressure Horner plot (Figure 3-18). Second, the errors in the slope are only ~15% on the average and appear to correlate inversely with the length of the injection period. This makes intuitive sense because the longer the injection period, the greater the likelihood of pressure and temperature changes reaching (quasi) stable values. This should be in agreement with the slope ratios obtained from the subsequent warmback/falloff periods.

3.7 Discussion and Concluding Remarks

The basic idea of this paper is to develop an approach for using temperature data to make inferences regarding the permeability (by using some information about the overall pressure response trends). We have been able to demonstrate that $\Delta T/\Delta P$ ratio from quasi-stabilized injection sequence responses is related to thermal properties and the permeability-thickness (kh) product. However, the relationship between q_h and q_f is not definitive (because of a normalizing constant that does not appear have a physical meaning). As noted earlier, the calculated value of this constant does not agree with the observed temperature difference – so we have proposed a workflow to bypass the direct use of this quantity. In this new approach, one needs the $\Delta T/\Delta P$ ratio and at least one warmback analysis to estimate the slope of the pressure Horner plot and hence the kh. At the very least, this would provide an independent estimate of k, especially if the transient pressure analysis turns out to be inconclusive.

Some of the complicating factors in this approach include:

- **Variable surface injection temperatures** – which would affect the temperature at which the injected CO_2 enters into the formation, and thus influence the BHT response in a complicated and non-linear manner. However, as long as sandface conditions are relatively stable so that phase changes do not occur during the injection period, the assumption that the CO2 temperature at the sandface is constant (and hence the heat flux associated with fluid injection) can be made. In the present study, the surface injection temperature was 68.7±11.7 oF (20.4±6.5 oC) during the period of investigation, so it can be considered as reasonably stable.
- **Short warmback periods between injection** – which can affect the span of the data available for the Horner straight-line analysis, and thus potentially compromise the robustness of the pressure Horner plot slope and permeability estimates. In the present study the duration of the warmbacks of interest range from 7.4 to 33.5 days, which appears to be adequate as evidenced by fairly long linear segments in the Horner plots.

- **Phase change of injected CO_2 –** if the injected CO_2 undergoes phase changes because of varying P and T conditions (e.g., from supercritical to liquid and vice versa) resulting in a step change in fluid-related thermal properties, it would invalidate many of the assumptions in the simple analytical models for the transient temperature analysis. In the present study, the CO_2 was always at supercritical conditions at the sandface.
- **Reservoir permeability changing with time** – should permeability change because of precipitation, wellbore plugging, etc., the analysis procedure would become more complicated and require the addition of a “skin zone” in the interpretive equations. However, in the present study, the consistency of the rate-normalized Horner plots strongly suggests that permeability did not change over the sequences of interest.

The new workflow for analyzing BHT data as developed and demonstrated here can be summarized as follows:

1. Characterize the injection sequence response in terms of quasi-steady temperature and pressure changes, DTQSS and DPQSS, respectively.
2. Analyze the following warmback response in terms of a temperature Horner plot. If multiple sequences are available, these can be analyzed together in terms of a rate-normalized temperature Horner plot as per Eqn. 3-3.
3. From the slope of this plot, and the ratio from the quasi-steady temperature and pressure changes as obtained from step (1), the slope of a corresponding Horner plot (or a rate-normalized pressure Horner plot) can be determined using Eqn. 3-9.
4. Subsequently, the permeability-thickness product can be estimated using Eqn. 3-8.

In conclusion, the fundamental contribution of this study is a new approach for analyzing BHT data that combines the amplitude of pressure and temperature changes from the injection period, with the temperature Horner analysis of one or more subsequent warmback periods, to estimate the slope of an equivalent pressure Horner plot, and hence, the permeability of the reservoir.

3.8 References

1. Abbaszadeh M., 2009, "Injection well testing," In: Kamal, MM, editor, Transient well testing, Richardson, TX, Society of Petroleum Engineers; p. 654-728.

2. Ambastha AK, 1989, "Pressure Transient Analysis for Composite Systems, PhD dissertation, Stanford University, Stanford, CA.

3. App, J., and K. Yoshioka, 2013, "Impact of Reservoir Permeability on Flowing Sandface Temperatures: Dimensionless Analysis," Society of Petroleum Engineers. doi:10.2118/146951-PA.

4. Benson SM, C. Doughty, 2006, "Estimation of field-scale relative permeability from pressure transient tests. In: Proceedings, EPA CO2SC Workshop, Berkeley, CA: Lawrence Berkeley National Laboratory.

5. Deflandre, J.P., A. Estublier, A. Baroni, J.-M. Daniel, and F. Adjémian, 2011, "In Salah CO_2 injection modeling: A preliminary approach to predict short term reservoir behavior," Energy Procedia, 4:3574–3581.

6. DOE/NETL, 2017, BEST PRACTICES: Monitoring, Verification, and Accounting (MVA) for Geologic Storage Projects, DOE/NETL-2017/1847, accessed at https://www.netl.doe.gov/sites/default/files/2018-10/BPM-MVA-2012.pdf.

7. Doughty, C., B. Freifeld and R.C. Trautz, 2002, "Site characterization for CO_2 geologic storage and vice versa: The Frio brine pilot, Texas, USA, as a case study," Mathematical Geology, 34(5):573-596.

8. Dowdle, W. L., and W.M. Cobb, 1975, "Static Formation Temperature From Well Logs - An Empirical Method," Society of Petroleum Engineers. doi:10.2118/5036-PA.

9. Eppelbaum, V. and I.M. Kutasov, 2006, "Pressure and temperature drawdown well testing: similarities and differences," J. Geophys. Eng., 3:12–20 doi:10.1088/1742-2132/3/1/002.

10. Gouturbe, B., F. Lucazeau, and A. Bonneville, 2007, "Comparison of several BHT correction methods: a case study on an Australian dataset," Geophys. J. Int., 170:913–922 doi: 10.1111/j.1365-246X.2007.03403.x.

11. Gupta, N., M. Kelley, R. Osborne, M. Moody, J. Gerst, S. Mishra, E. Howat, C. Sullivan, G. Spitznogle, I. Bhattacharya and M. Hammond, 2012, "Development of conceptual design for commercial scale geologic storage and monitoring system at American Electric Power Mountaineer plant," Proc., Intl Conf on Greenhouse Gas Technologies (GHGT-11), Kyoto, Japan, Nov. 18-22.

12. Gupta N, Jagucki P, Sminchak JR, Kelley ME, Spitznogle, G, Sherrick, B., 2009, "Integration and deployment of geologic storage and monitoring across two deep saline reservoirs with post-combustion capture at American Electric Power Mountaineer Plant," Proc. 9th Intl. Conf Greenhouse Gas Technologies (GHGT-9), Amsterdam, Netherlands.

13. Jarrell, P.M., C. Fox, M. Stein and S. Webb, 2002, Practical Aspects of CO_2 Flooding, SPE Monograph 22, Society of Petroleum Engineers, Richardson, TX, 214 pp.

14. Kempka, T., M. Kühn, 2013, "Numerical simulations of CO_2 arrival times and reservoir pressure coincide with observations from the Ketzin pilot site, Germany," Environmental Earth Sciences, 70:3675 –3685.

15. Kutasov, I.M. and V. Eppelbaum, 2005, "Determination of formation temperature from bottom-hole temperature logs—a generalized Horner method," J. Geophys. Eng. 2:90–96 doi:10.1088/1742-2132/2/2/002.

16. Mao, Y., and M. Zeidouni, 2017, "Analytical Solutions for Temperature Transient Analysis and Near Wellbore Damaged Zone Characterization," Society of Petroleum Engineers. doi:10.2118/185990-MS.

17. Mawalkar, S., D. Brock, A. Burchwell, M.Kelley, S. Mishra, N. Gupta, R. Pardini and B. Shroyer, 2019, "Where is that CO_2 flowing? Using DTS technology for monitoring injection of CO_2 into a depleted oil reservoir," Intl. J. of Greenhouse Gas Control, 85:132-142.

18. Mishra, S., M. Kelley, E. Zeller, N. Slee, N. Gupta, I. Bhattacharya and M. Hammond, 2012, "Maximizing the value of pressure monitoring data from CO_2 sequestration projects," Proc., International Conference on Greenhouse Gas Technologies (GHGT-11), Kyoto, Japan, Nov. 18-22.

19. Mishra, S., Y. Oruganti, N. Gupta, P. Ravi Ganesh, I. Bhattacharya and G. Spitznogle, 2014, "Modeling CO_2 plume migration based on calibration of injection and post-injection pressure response at the AEP Mountaineer Project," Greenhouse Gasses: Science & Technology, 4:331-356.

20. Muradov, K. and D. Davies, 2012, "Temperature transient analysis in horizontal wells: Application workflow, problems and advantages," Journal of Petroleum Science and Engineering, 92-93:11–23.

21. Nunez-Lopez, V., Munoz-Torres, J., Zeidouni, M., 2014. Temperature monitoring using distributed temperature sensing (DTS) technology. Energy Procedia 63, 3984–3991.

22. Onur, M., and M. Cinar, 2017, "Modeling and Analysis of Temperature Transient Sandface and Wellbore Temperature Data from Variable Rate Well Test Data," Society of Petroleum Engineers. doi:10.2118/185802-MS.

23. Ramey, HJ Jr., 1970, “Approximate solutions for unsteady liquid flow in composite reservoirs,” J Cdn Pet Tech, 32-37.

24. Robertson, E.G. 1988, Thermal Properties of Rocks, USGS Open File Report 88-141, United States Geological Survey, Reston, VA, 106 pp.

25. Sato, T., S. White and Z. Xue, 2006, “Numerical modeling of Nagaoka injection test site in Nigata, Japan,“ Proc., TOUGH Symposium 2006, Lawrence Berkeley National Laboratory, Berkeley, CA, May 15-17.

26. Singh, V., A. Cavanagh, H. Hansen, B.Nazarina, M. Idling and P. Ringrose, 2010, “Reservoir modeling of CO_2 plume behavior calibrated against monitoring data from Sleipner, Norway,” paper SPE 134891 presented at SPE Annual Technical Conference and Exhibition, Florence, Italy, Sept. 19-22.

27. Sminchak J, Gupta N, Gerst JL, 2009, “Well test results and reservoir performance for a carbon dioxide injection test in the Bass Islands Dolomite in the Michigan Basin,” Env Geosci, 16(3):153-162.

Chapter 4.

Injectivity Index – A Powerful Tool for Characterizing CO_2 Storage Reservoirs

Manoj Valluri, Srikanta Mishra, and Priya Ravi Ganesh

Chapter 4. Table of Contents

Chapter 4. List of Tables

Chapter 4. List of Figures

Chapter 4. List of Attachments

4.1 Motivation and Scope

In oil and gas operations, productivity (or injectivity) index is a popular concept for evaluating the capability of a well to produce from (or inject fluids into) a porous and permeable formation [22]. In the context of CO_2 injection into saline aquifers for geologic storage (sequestration), it is useful to focus on the injectivity index which is defined as the ratio of the injection rate divided by the pressure difference between formation pressure and bottom-hole pressure:

$$J = \left|\frac{q}{(P_f - P_{BH})}\right| \qquad \textbf{(Eqn. 4-1)}$$

where J is injectivity index, q is injection rate, P_f is reference formation pressure and P_{BH} is bottom-hole pressure. During a typical CO_2 injection event, the injection rate, q, is maintained at a relatively constant rate, and the bottom-hole pressure rapidly increases to some equilibrium value, P_{BH}, after which it changes slowly. The difference between this quasi-equilibrium value and some reference formation pressure, P_f (i.e., stable pressure prior to injection), is the denominator in Eqn. 4-1.

From a theoretical standpoint, the injectivity index can be related to the permeability-thickness product of the formation and the size of the reservoir [22]. From a practical standpoint, the injectivity index helps compare the relative potential of different formations for injection operations corresponding to a prescribed pressure differential. It is also useful for determining the change in well performance over time.

The first application of the injectivity index concept in the Carbon Capture and Storage (CCS) literature appears to be that of Mishra [14], who evaluated field injection pressure and rate data from the AEP Mountaineer project for two different formations. A number of discrete injection events were analyzed, and a corresponding range of injectivity index values were calculated for each formation. In general, these values were found to be strongly correlated with independent assessments of the permeability-thickness product.

Ravi Ganesh and Mishra [18] presented a correlating chart between injectivity index (expressed in MT/yr/psi) and permeability thickness (expressed in mD-ft) using data from two additional field projects as well as a number of numerical simulation runs for various parameter combinations. They found a strong trend over multiple orders of magnitude and suggested the relationship $J \sim 0.1\ kh$ as a rough order-of-magnitude type approximation. Mishra *et al.*, [15] further confirmed the validity of this trend using synthetic (i.e., simulated) data for CO_2 injection into depleted oil reservoirs. They concluded that the trend is well described by $J \sim 0.04\ kh$ at the lower bound and $J \sim 0.14\ kh$ at the upper bound – with $J \sim 0.1\ kh$ appearing to be adequate for scoping calculations.

The objective of this technical note is to present field data from multiple additional CCS projects to corroborate the findings summarized above, and further demonstrate the utility of the injectivity index as a reservoir characterization tool via its strong correlation with the underlying permeability-thickness value.

4.2 Description of Field Data

Field data was collected for nine global CCS projects listed in Table 4-4. The projects used for this study spanned across various continents and injection scales categorized as pilot (~10,000 MT/year), demonstration (~1 million MT/year), and commercial (>1 million MT/year).

Table 4-4. CCS Projects in this Study.

Ketzin **[13, 21])**	Brandenburg, Germany	Pilot
Illinois Basin Decatur Project (IBDP) **[21, 12])**	Decatur, Illinois, USA	Demonstration
Snohvit **[4, 20, 3]**	Barents Sea, Norway	Commercial
Nagaoka **[13, 17]**	Nagaoka, Japan	Pilot
SECARB **[11, 13]**	Kemper County, Mississippi, USA	Demonstration
Aquistore **[19, 23, 5, 7]**	Saskatchewan, Canada	Demonstration
Midwest Regional Carbon Sequestration Partnership (MRCSP) **[1, 2]**	Otsego County, Michigan, USA	Pilot
	Rabbit Hash, Kentucky, USA	Pilot
AEP Mountaineer **[14, 10]**	New Haven, West Virginia, USA	Pilot

Figure 4-19 is a map showing the geographic locations of the nine projects.

Projects are numbers as: 1 – Aquistore, Canada, 2 – MRCSP (Michigan, USA), 3 – AEP Mountaineer Project (West Virginia, USA), 4 – IBDP (Illinois, USA), 5 – MRCSP (Kentucky, USA), 6 – SECARB (Mississippi, USA), 7 – Snohvit (Norway), 8 – Ketzin (Germany), and 9 – Nagaoka (Japan).

Figure 4-19. Map showing the global CCS projects used for data population in this study.

Table 4-5 presents a summary of basic geologic and operational information from these projects, including formation type, depth of injection zone, average porosity, injection and monitoring well names, and average injection rate.

Table 4-5. Summary Information for CCS Projects in this Study.

Ketzin	Stuttgart sandstone	650 (2,133)	26%	Ktzi 201	Ktzi 200, Ktzi 202	27 (0.9)	50
IBDP	Mount Simon sandstone	2,120 (6,988)	20%	CCS1	VW1, GM1	504 (17.8)	936
Snohvit	Tubåen sandstone	2,600 (8,530)	7-20%	F-2H	F-2H	874 (30.9)	1,625
Nagaoka	Haizume sandstone	1,100 (3,609)	22.5%	IW-1	OB-2, OB-3, OB-4	19 (0.7)	35
SECARB	Tuscaloosa sandstone	3,117 (10,300)	20-25%	29-10, 29-12, 25-2, 24-2, 29-2, 48-1, 29-7, 26-1, 27-1, 28-1, 29-4	EGL7	1,342 (47.4)	2,495
Aquistore	Black Island and Deadwood sandstone	2,195 (7,200)	11-17%	PRTC-Inj-5626	PRTC-Obs-5628	177 (6.2)	329
MRCSP - Michigan	Bass Islands dolomite	1,049 (3,442)	13%	Charlton 4-30	Charlton 3-30A	282 (10.0)	524
MRCSP - East Bend	Mount Simon sandstone	985 (3,230)	5-15%	East Bend well	East Bend well	427 (15.1)	794
AEP Mountaineer - Copper Ridge	Copper Ridge dolomite	2,482 (8,144)	5-15%	AEP-1	MW-2	117 (4.1)	218
AEP Mountaineer - Rose Run	Rose Run sandstone	2,362 (7,749)	5-12%	AEP-2	MW-3	32 (1.1)	60

The data used for analyzing injectivity is presented in the Appendix in the form of injection rate and downhole pressure versus time charts for each of the projects. In these plots, the red boxes identify discrete injection events under the same injection activity that were isolated and analyzed separately in this study to obtain injectivity index estimates. The plots also show a baseline pressure which was used as a reference pressure to calculate the pressure buildup corresponding to an injection event.

4.2.1 Reservoir Properties

Table 4-6 shows basic reservoir properties (extracted from the citations in Table **4-4**) for the injection zones identified in the nine CCS projects in SI units. Values in parentheses represent field units. CO_2 formation volume factors were estimated at the appropriate pressure and temperature values using NIST tables [6]. Brine viscosity was determined using the correlations presented in McCain [9].

Table 4-6. Reservoir Properties for the CCS Projects in this Study.

P_R – reservoir pressure, T_R – reservoir temperature, S – formation brine salinity, k – reservoir absolute permeability, h – injection zone thickness, B_{CO2} – CO_2 formation volume factor, μ_w – formation brine viscosity.

Project Name	**P_R, kPa (psi)**	**T_R, °C (°F)**	**S, ppm**	**kh, m^3 (mD-ft)**	**CO_2 formation volume factor, res. ft^3/ft^3**	**μ_w, Pa.s (cP)**
Ketzin	6,205 (900)	33 (91)	250,000	5.52E-13 (1,835)	0.0106	0.0015 (1.45)
IBDP	23,001 (3,336)	63 (146)	110,000	3.79E-12 (12,600)	0.0025	0.0006 (0.64)
Snohvit	28,503 (4,134)	95 (203)	100,000	7.77E-12 (25,840)	0.0028	0.0005 (0.47)
Nagaoka	11,997 (1,740)	47 (117)	7,113	8.42E-14 (280)	0.0030	0.0007 (0.68)
SECARB	11,721 (1,700)	43 (110)	200,000	1.77E-11 (59,000)	0.0028	0.0006 (0.64)
Aquistore	34,143 (4,952)	104 (220)	250,000	3.69E-13 (1,227)	0.0027	0.0007 (0.65)
MRCSP - Michigan	10,687 (1,550)	28 (83)	350,000	1.10E-12 (3,650)	0.0023	0.0016 (1.57)
MRCSP - East Bend	11,032 (1,600)	27 (80)	203,000	2.47E-12 (8,200)	0.0023	0.0014 (1.44)
AEP Mountaineer - Copper Ridge	27,110 (3,932)	63 (145)	300,000	6.92E-12 (23,000)	0.0023	0.0012 (1.21)
AEP Mountaineer - Rose Run	25,262 (3,664)	54 (129)	300,000	8.43E-14 (280)	0.0023	0.0012 (1.19)

4.3 Calculation Approach

Injectivity index, *J*, was calculated for all the nine CCS projects discussed above. Injection rate and/or cumulative injection data was first examined to divide the data into discrete injection events. Each of these events, *i*, consisted of a finite amount of CO_2, Q_i injected at one specific rate or a stepwise series of rates over a period Δt_i. This led to a corresponding downhole pressure increase, ΔP_i – calculated as the difference between the maximum pressure attained during an injection event, P_{BH}, and a baseline (reference) pressure, P_f. Consequently, the event specific *J* value, J_i was calculated as:

$$J_i = \frac{Q_i}{\Delta P_i * \Delta t_i} \quad or \quad J_i = \frac{q_i}{\Delta P_i} \qquad \textbf{(Eqn. 4-2)}$$

where q_i is the average CO2 injection rate,

J is typically expressed in the units of MT/yr-kPa or MT/yr-psi.

4.3.1 Calculated Injectivity Index Values and Updated Correlation

Table 4-7 lists the calculated injectivity index values in SI units. Values in braces denote injectivity index values in field units. Calculations over multiple events are presented as average, minimum, and maximum values.

Table 4-7. Injectivity Indices for the CCS Projects in this Study**.**

	J, MT/yr-kPa (MT/yr-psi)			Events
Project Name	**Average**	**Min**	**Max**	No.
Ketzin	23 (157)	16 (110)	223	4
IBDP	144 (990)	97 (671)	1,176	6
Snohvit	111 (766)	58 (403)	1,007	7
Nagaoka	8 (54)	6 (40)	62	4
SECARB	780 (5,379)	167 (1,153)	11,526	4
Aquistore	25 (169)	13 (87)	327	4
MRCSP - Michigan	54 (373)	32 (219)	438	4
MRCSP - East Bend	163 (1,123)	106 (730)	1,327	3
AEP Mountaineer - Copper Ridge	251 (1,731)	131 (900)	341 (2,349)	19
AEP Mountaineer - Rose Run	5 (37)	2 (11)	10 (68)	25

Figure 4-20 is a log-log plot of event-based average injectivity index (J_{avg}) versus *kh* for all the nine CCS projects discussed in this study. The yellow circles represent the value of *J* calculated for each individual injection event for these projects. The blue diamond represents the average *J* value. As noted earlier, Mishra *et al.*, [15] reported a strong correlation between injectivity index and *kh* and noted that the field data examined lied between two lines represented by *J* = 0.04 *kh* and *J* = 0.14 *kh* where *J* is given in [MT/yr-psi] and *kh* is given in [mD-ft]. These two bounds are shown as dotted lines in Figure 4-20. The new bounds obtained after incorporating data from the additional global CCS projects considered in this study are plotted as dashed lines in the same figure. These lines bracket the range of average *J* values obtained for various projects. They correspond to *J* = 0.03 *kh* (lower bound) and *J* = 0.23 *kh* (upper bound) and should be considered as the updated bounds to *J* v/s *kh* relationship. Finally, the median trend expressed by the relationship *J = 0.08 kh* can be used for single-point estimates. These results corroborate our earlier findings on the correlation between *J* and *kh*, while refining the nature of the correlation and its robustness by using data from prominent CCS projects across the globe.

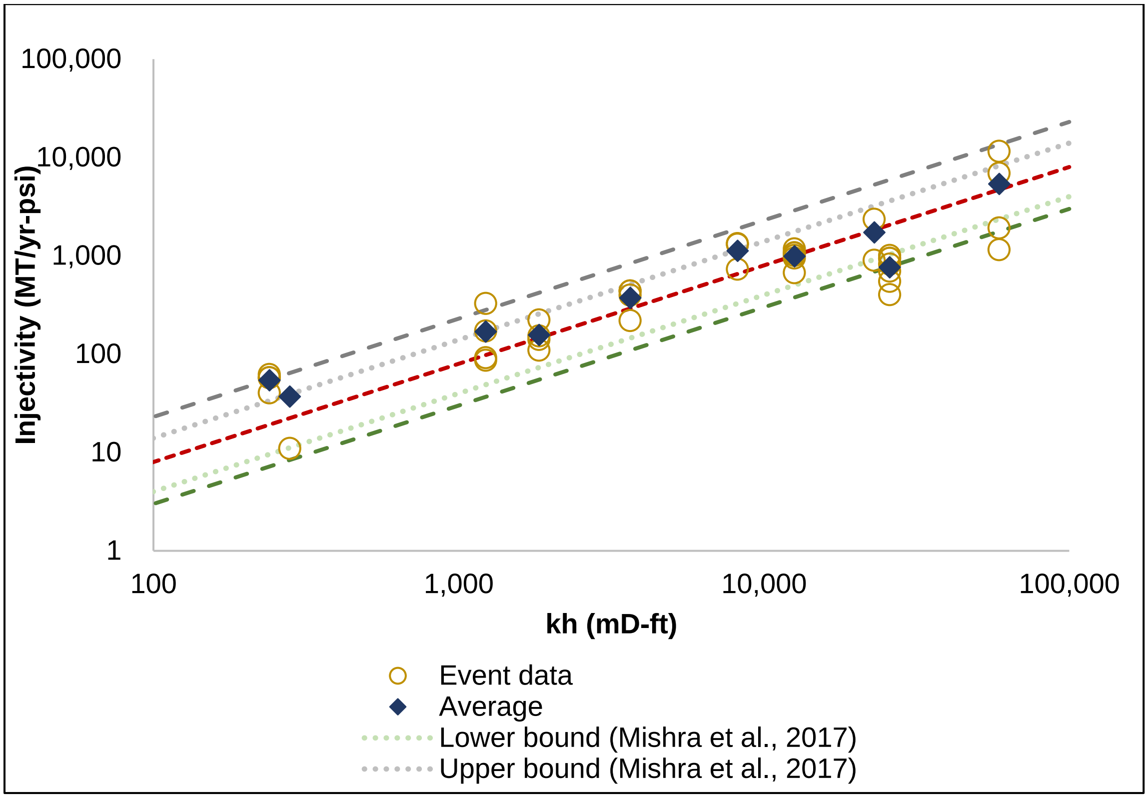


Figure 4-20. Estimated injectivity index values vs. permeability thickness (kh) product from current study, superposed on earlier results from Mishra *et al.* [15]

4.4 Discussion

In the context of CO_2 geological storage projects, we believe the utility of the injectivity index concept is twofold. First, it allows a rapid estimation of pressure buildup corresponding to a target injection rate. For example, if the target injection rate *q* for the IBDP project is 1 million MT/yr, and its injectivity index *J* is ~1000 MT/yr/psi, then the expected pressure buildup will be $\Delta P = q/J$ = 1 million /1000 = 1000 psi. This calculation can be readily inverted to yield the expected injection rate if there is a target pressure buildup (so as not to exceed fracture pressure related permit constraints). For example, if the maximum allowable pressure buildup is 800 psi, then the maximum rate should be 1000*800 = 800,000 MT/yr. In the planning stages of a project, the correlation in Figure 4-20 can be used to assist such calculations by pointing to analog projects and also by providing an uncertainty range.

The second benefit is in rapidly estimating the permeability-thickness product without undertaking any detailed injection-falloff analysis using standard transient pressure analysis techniques. Returning to the IBDP example, for an average injectivity index *J* = 1000 MT/yr/psi, the median *kh* value = 1000/0.08 = 12500 md-ft, which agrees quite well with the reported value of 12600 md-ft (Table 4-6). This shows the power of the correlation presented in Figure 4-20, which can also be used to assign uncertainty bounds to these calculations.

Another benefit of this concept is in assessments of field data where the injection rate has significant fluctuations over time (as can be seen in the plots in the Appendix for almost all of the projects investigated). In such conditions, the injectivity index concept allows the operator to track injection performance over time, and determine if there is any actual and systematic change in injectivity (either increase or decrease from physical effects), rather than pressure changes simply associated with rate variations.

In summary, we have presented field data from multiple CCS projects to corroborate earlier simulation studies regarding the strong correlation between injectivity index and the underlying permeability-thickness value. We hope that this concept will become a powerful tool for characterizing CO_2 storage reservoirs via: (a) ready calculation of expected pressure buildup given a target injection rate (and vice versa), and (b) providing independent permeability-thickness estimates.

4.5 References

1. Battelle, 2010 (a). CO_2 Injection Test in the Cambrian-Age Mt. Simon Formation Duke Energy East Bend Generating Station, Boone County, Kentucky. Report prepared for U.S. DOE-NETL under cooperative agreement number DE-FC26-05NT42589, December 2010.

2. Battelle, 2010 (b). Michigan Basin Phase II Geologic CO_2 Sequestration Field Test. Report prepared for U.S. DOE-NETL under cooperative agreement number DE-FC26-05NT42589, May 2010.

3. Grude, S., Landrø, M., and Dvorkin, J., 2014. Pressure effects caused by CO_2 injection in the Tubåen Fm., the Snøhvit field. International Journal of Greenhouse Gas Control 27 (2014), pp 178-187.

4. Hansen, O., *et al.*, 2013. Snøhvit: The history of injecting and storing 1 Mt CO_2 in the fluvial Tubåen Fm. Energy Procedia 37 (2013), pp 3565-3573.

5. Houseworth *et al.*, 2011. Leakage Risk Assessment for a Potential CO_2 Storage Project in Saskatchewan, Canada. Report LBNL-4915E published May 11, 2011.

6. Jarrell, P.M., Fox, C.E., Stein, M.H., Webb, S.L. 2002. Practical Aspects of CO_2 Flooding. SPE Monograph Series. Society of Petroleum Engineers, v. 22, 220 p.

7. Jiang *et al*., 2016. Geologic Modeling and Simulation Report for the Aquistore Project. Report 2016-EERC-04-06 prepared for U.S. DOE-NETL under cooperative agreement number DE-FC26-05NT42592, February 2016.

8. Liebscher, A. *et al*., 2013. Injection operation and operational pressure-temperature monitoring at the CO_2 storage pilot site, Ketzin, Germany – Design, results, and recommendations. International Journal of Greenhouse Gas Control 15 (2013), pp 163-173.

9. McCain, W. J.. 1989. "The Properties of Petroleum Fluids," Tulsa, OK, PennWell Books.

10. McNeil, C., *et al*., 2014. Lessons learned from the post-injection site care program at the American Electric Power Mountaineer Product Validation Facility. Energy Procedia 63 (2014), pp 6141–6155.

11. Meckel, T.A., and Hovorka, S.D., 2009. Results from Continuous Downhole Monitoring (PDG) at a Field-Scale CO_2 Demonstration Project, Cranfield, MS. Paper SPE 127087 presented at the 2009 SPE International Conference on CO_2 Capture, Storage, and Utilization held at San Diego, California, USA, November 2-4.

12. Mehnert, E., and Weberling, P.H., 2014. Groundwater Salinity Within the Mt. Simon Sandstone in Illinois and Indiana. Illinois State Geological Survey, Circular 582.

13. Michael, K. *et al*., 2010. Geological storage of CO_2 in saline aquifers – A review of the experience from existing storage operations. International Journal of Greenhouse Gas Control 4 (2010), pp 659-667.

14. Mishra, S., *et al*., 2013. Maximizing the Value of Pressure Monitoring Data from CO_2 Sequestration Projects. Energy Procedia 114 (2017), pp 3465–3475.

15. Mishra, S., *et al*., 2017. Analyzing the Performance of Closed Reservoirs Following CO_2 Injection in CCUS Projects. Energy Procedia 37 (2013), pp 4155–4165.

16. Mito, S., Xue, Z., and Ohsumi, T., 2008. Case study of geochemical reactions at the Nagaoka CO_2 injection site, Japan. International Journal of Greenhouse Gas Control 2 (2008), pp 309-318.

17. Mito, S., and Xue, Z., 2011. Post-Injection Monitoring of Stored CO_2 at the Nagaoka Pilot Site: 5 Years Time-Lapse Well Logging Results. Energy Procedia 4 (2011), pp 3284-3289.

18. Ravi Ganesh, P. and Mishra, S. 2014. Reduced physics models of CO2 injectivity. Energy Procedia (2014), Vol. 63, pp. 3116-3125.

19. Rostron *et al*., 2002. Economic Potential of Formation Brines: Interim Results from the Saskatchewan Brine Sampling Program. Saskatchewan

Geological Survey, Volume 2, Misc. Rep. 2002-4.2, CD-ROM, Paper C-1, 29p.

20. Shi, J. *et al.*, 2013. Snøhvit CO_2 storage project: Assessment of CO_2 injection performance through history matching of the injection well pressure over a 32-months period. Energy Procedia 37 (2013), pp 3267-3274.

21. Striebel, M. *et al*., 2014. From Pilot to Demo Scale – Comparing Ketzin Results with the Illinois Basin Decatur Project. Energy Procedia 63 (2014), pp 6323-6334.

22. Terry, R.E and Rogers, J.B. 2014. Applied Petroleum reservoir Engineering 3rd edition. Upper Saddle River, N.J., Prentice Hall, 528 p.

23. Whitaker, S. and Worth, K., 2011. Aquistore: a fully integrated demonstration of the capture, transportation and geologic storage of CO_2. Energy Procedia 4 (2011), pp 5607–5614.

Appendix A. – Injection Pressure and Rate History

Ketzin CCS Project, Germany

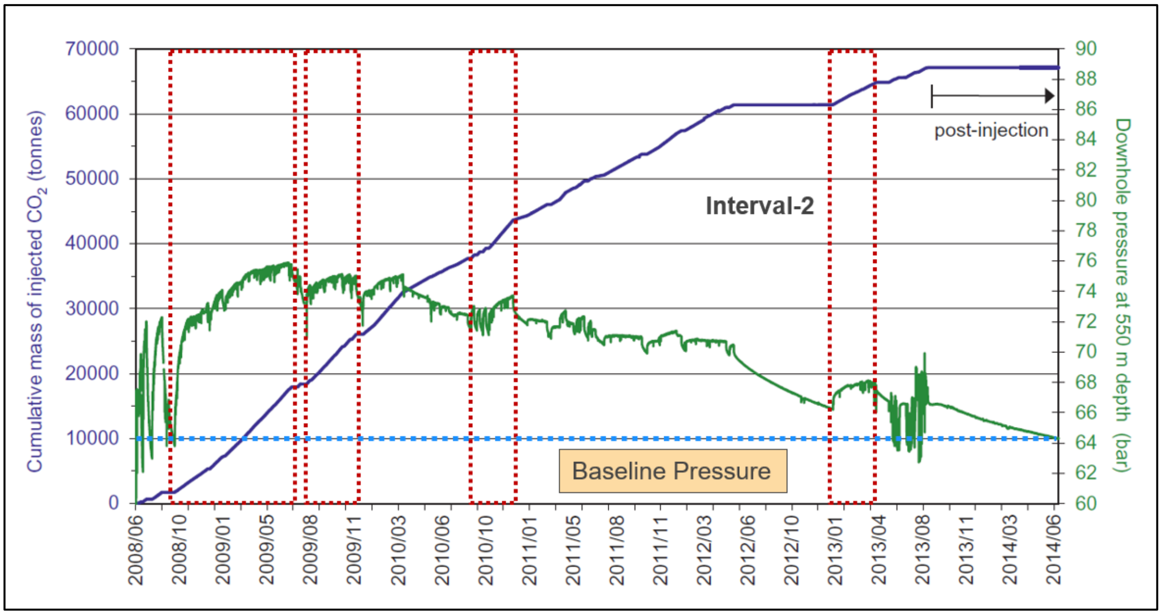


Figure A-1. Ketzin injection data; blue line represents cumulative CO_2 injected and the green line represents downhole pressure measurements at a depth of 1,805 feet (Striebel *et al.*, [21]).

Illinois Basin Decatur Project (IBDP), United States

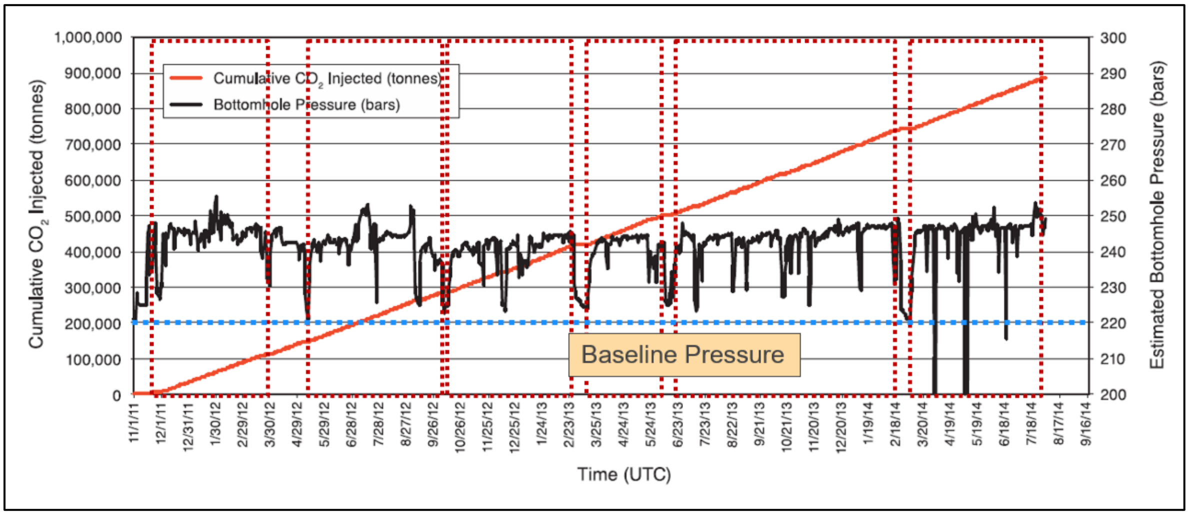


Figure A-2. IBDP injection data; red line represents cumulative CO2 injected and the black line represents downhole pressure estimation at the injector (Striebel *et al.*, [21]).

Snøhvit CCS Project, Norway

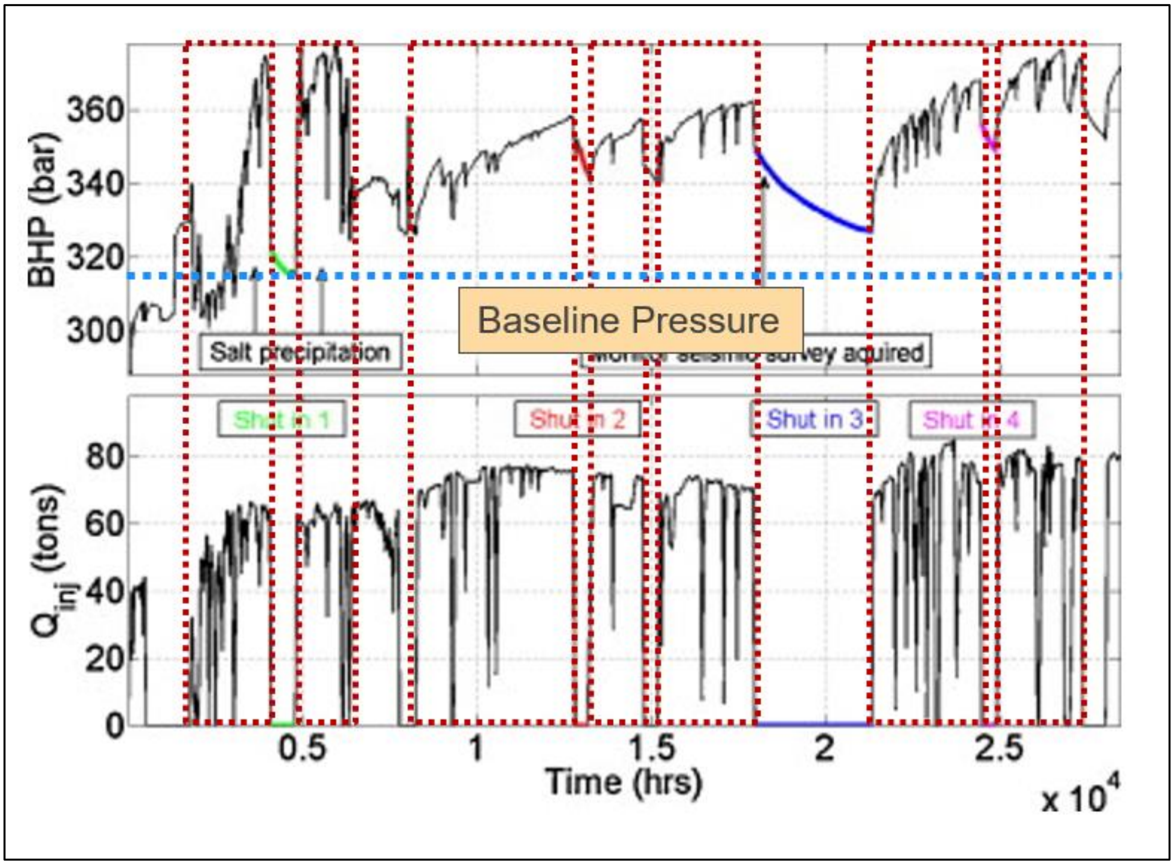


Figure A-3. Snøhvit CCS project operational data; Top – F-2H injector bottomhole pressure; Bottom – CO2 hourly injection rate (Grude *et al*., [3]).

Nagaoka CCS Project, Japan

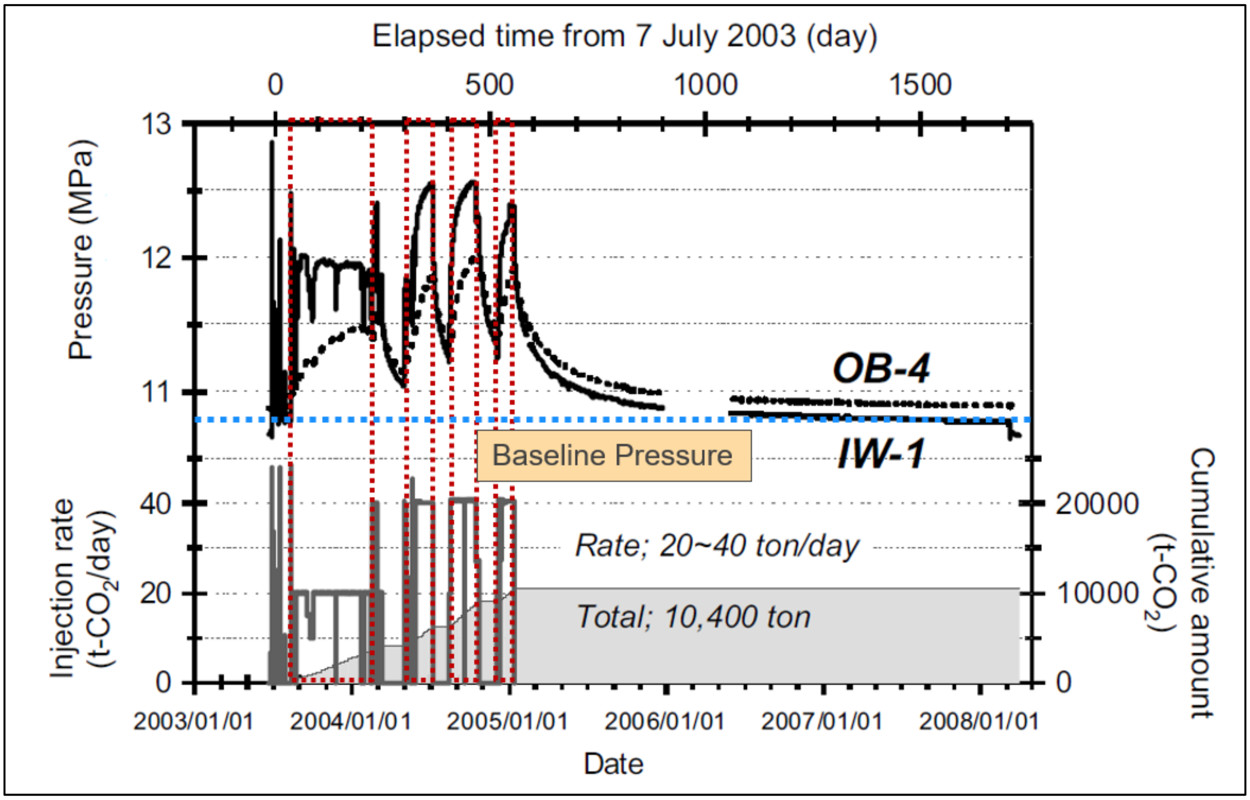
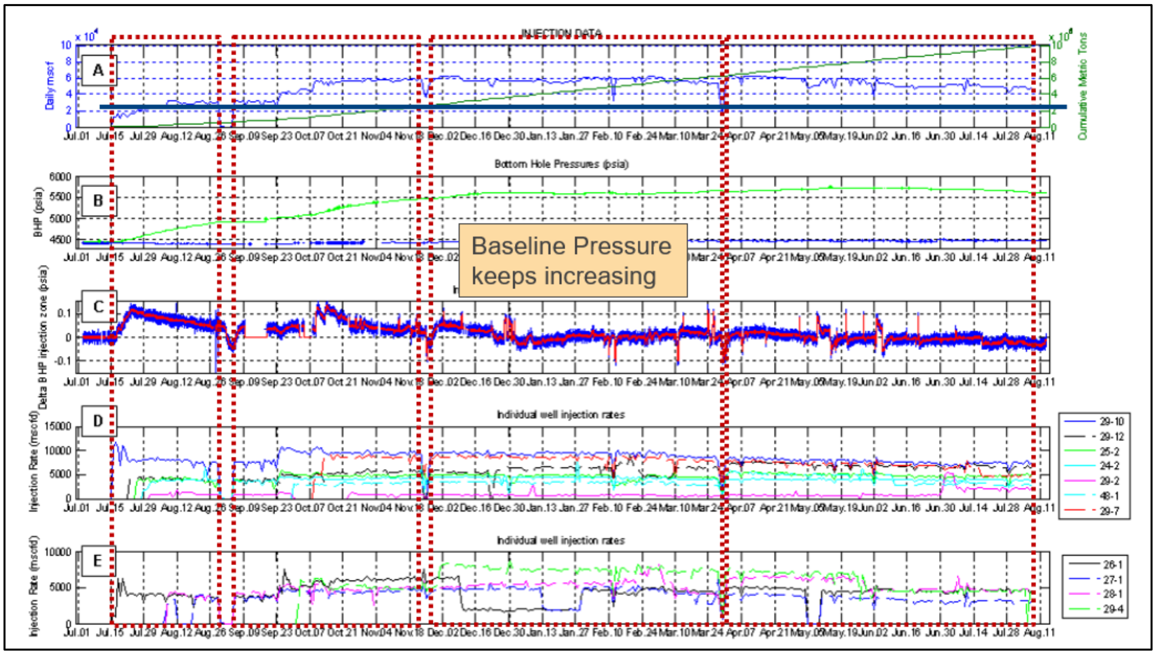


Figure A-4. Nagaoka CCS project operational data; Top – IW-1 and OB-4 bottomhole pressures; Bottom – CO2 daily injection rate (Mito *et al.*, [16]).

Southeast Regional Carbon Sequestration Partnership (SECARB), United States



Baseline Pressure keeps increasing

Figure A-5. SECARB Cranfield project operational data; A daily and cumulative injection data, B – Pressure response in injection zone and overlying monitoring zone, C – rate of pressure change in the injection zone, D – individual injection rates for injectors on the same side of the fault as the observation well, E – individual injection rates for injectors on the opposite side of the fault as the injector well,

Aquistore, Canada

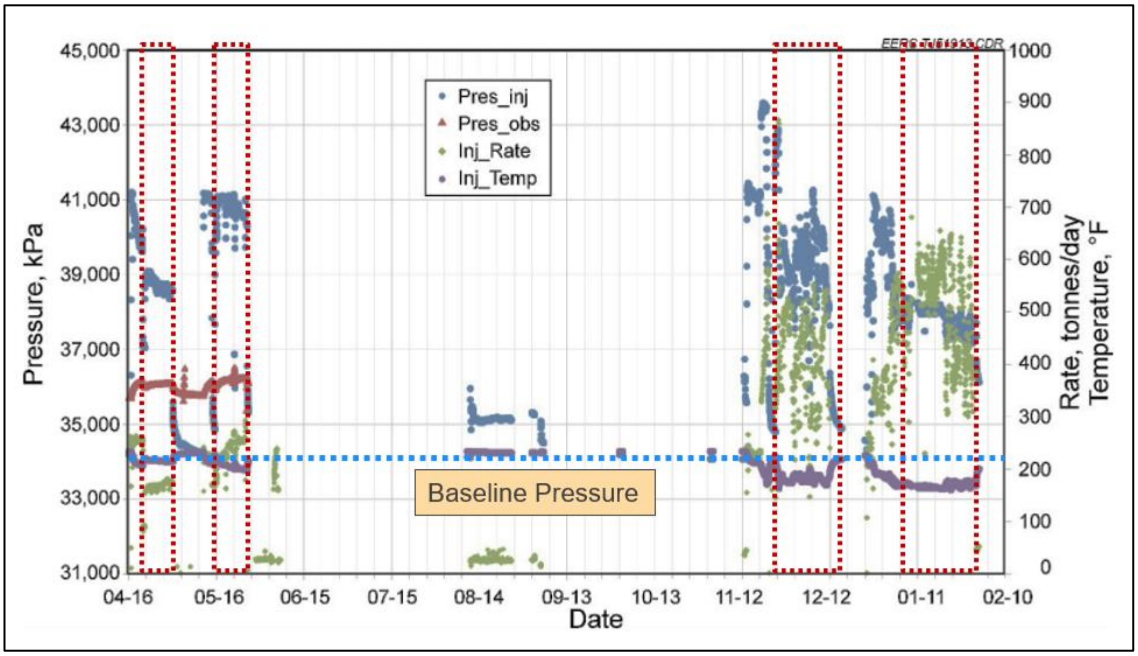


Figure A-6. Aquistore injection data (Jiang *et al*., [7])

Midwest Regional Carbon Sequestration Partnership (MRCSP), United States

Otsego County, Michigan

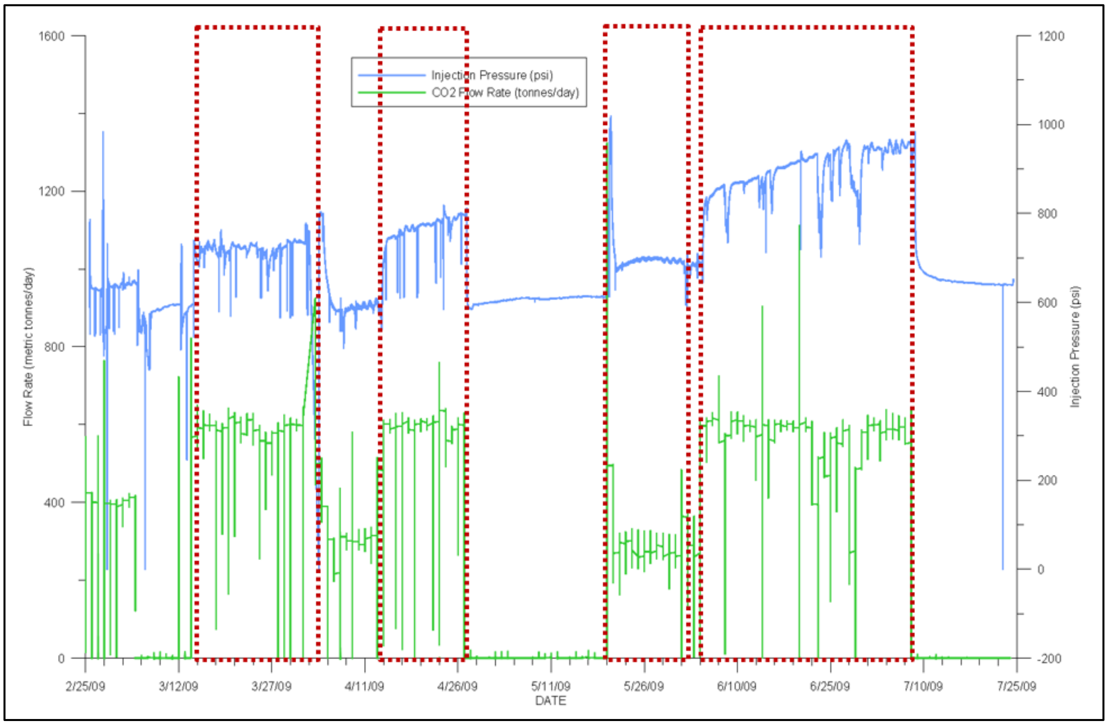


Figure A-7. Injection rate data from the Charlton 4-30 well (green lines) [1].

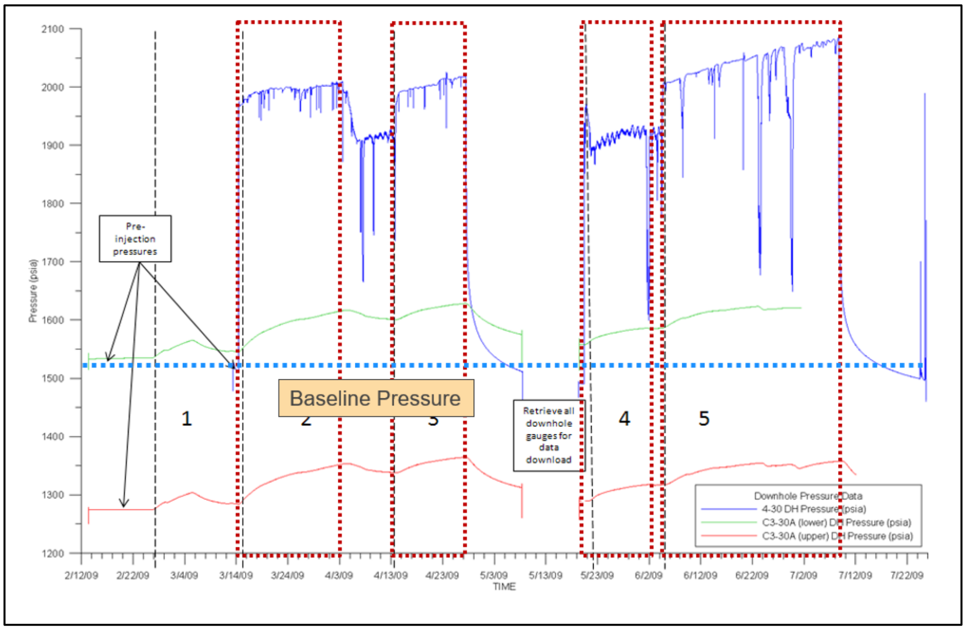


Figure A-8. Downhole pressure data from the Charlton 4-30 well (blue line) (Battelle [1]).

Boone County, Kentucky

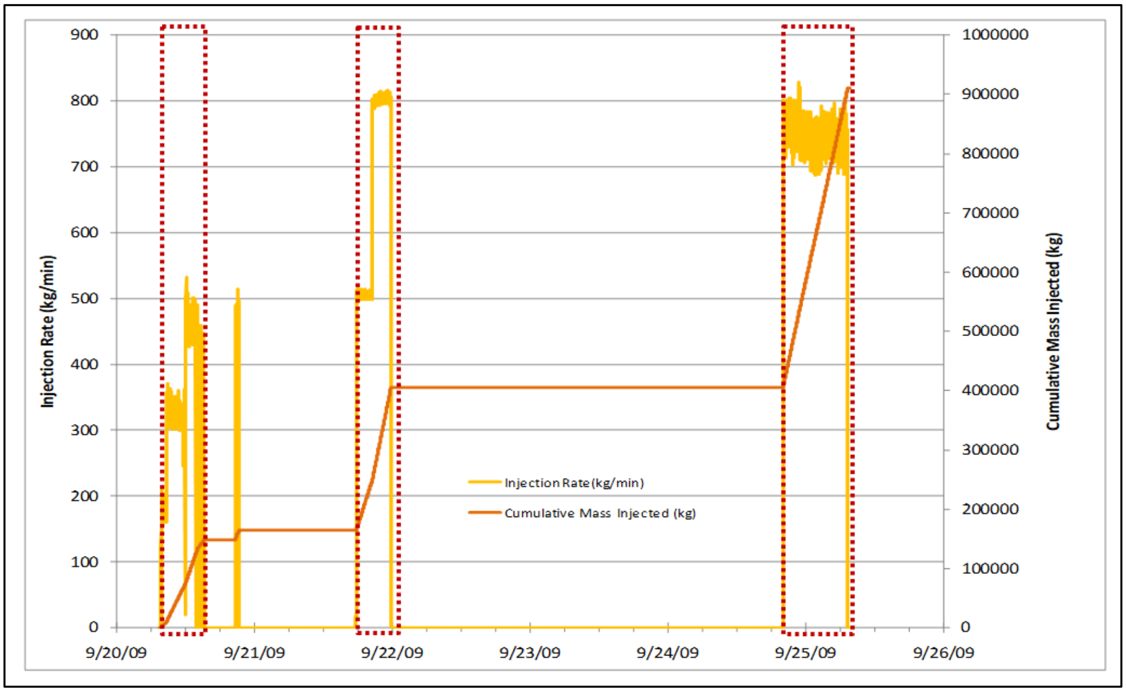


Figure A-9. Injection rate data for the East Bend site (Battelle [2]).

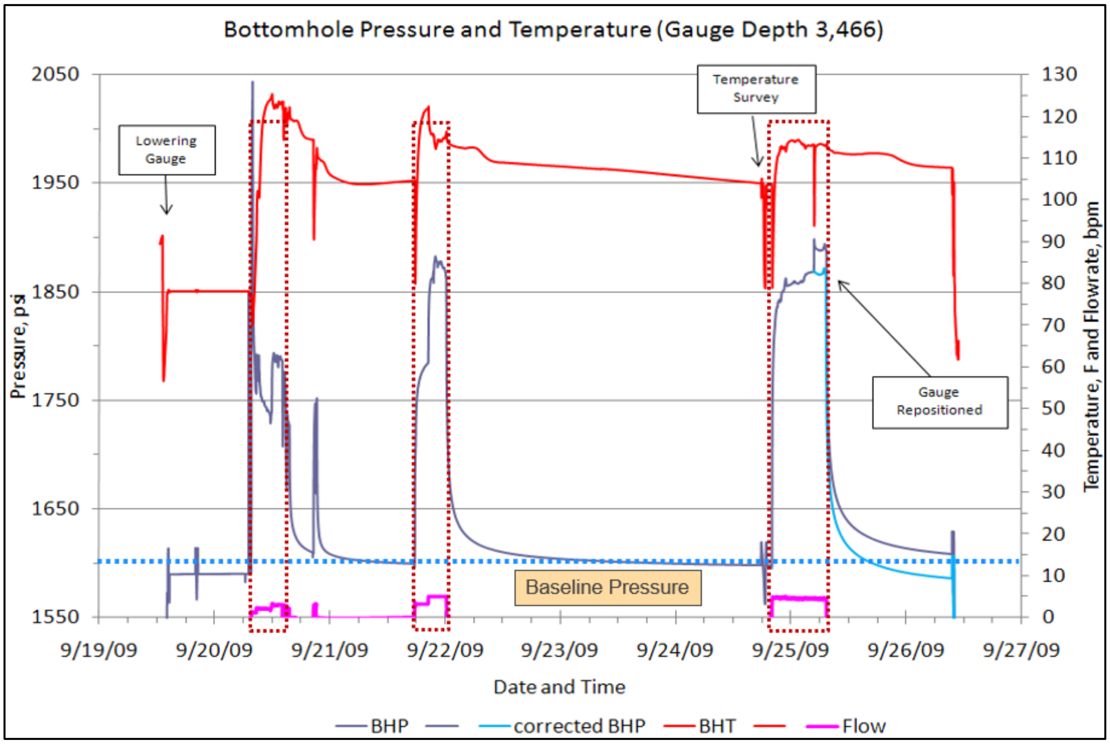


Figure A-10. Bottomhole pressure data for the East Bend site (Battelle [2]).

American Electric Power (AEP) Mountaineer CCS Project

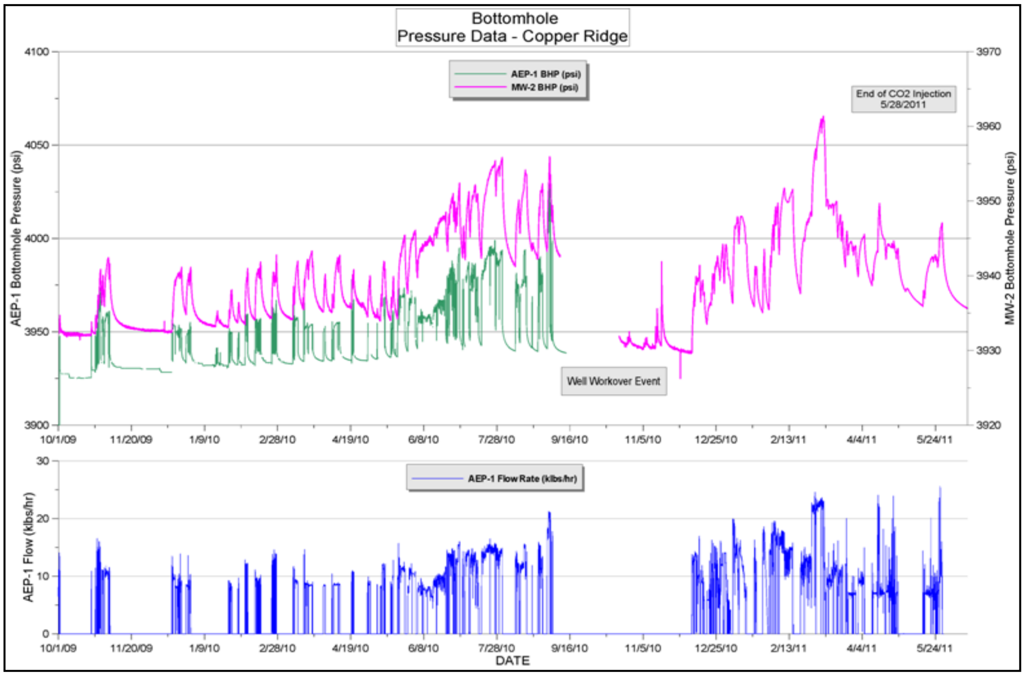


Figure A-11. Injection rate data for the Copper Ridge at the AEP Mountaineer site (Mishra *et al.*, [14]); green line – AEP-1 BHP, pink line – MW-1 BHP, blue line – AEP-1 injection rate.

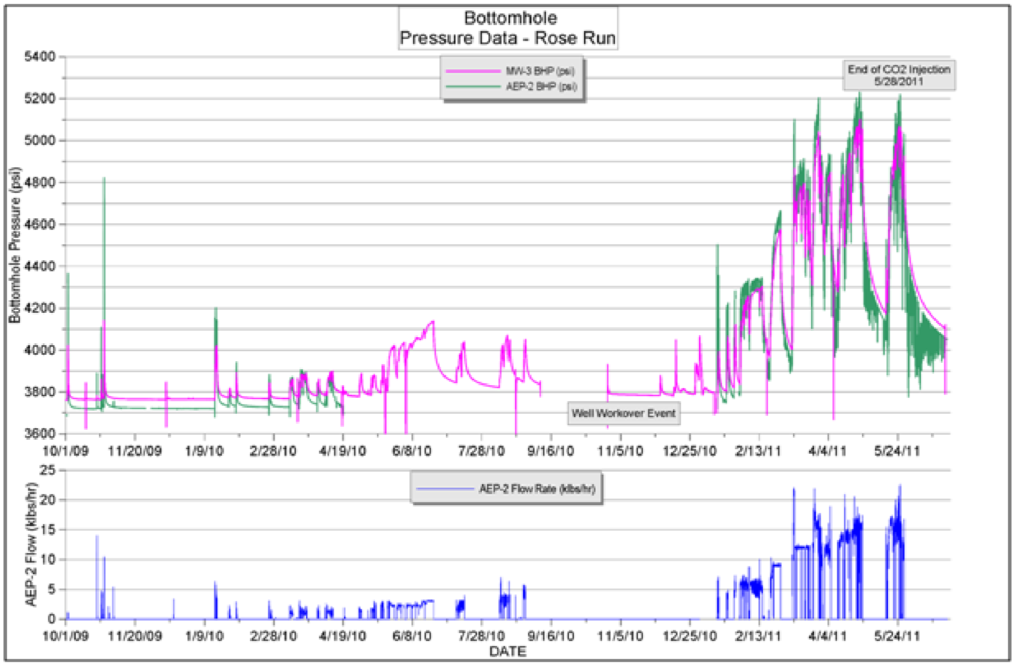


Figure A-12. Injection rate data for the Rose Run at the AEP Mountaineer site (Mishra *et al.*, [14]); green line – AEP-2 BHP, pink line – MW-2 BHP, blue line – AEP-2 injection rate.

Chapter 5.

A Screening Model for Predicting Injection Well Pressure Buildup and Plume Extent in CO_2 Geologic Storage Projects

Srikanta Mishra and Priya Ravi Ganesh

Chapter 5. Table of Contents

Chapter 5. List of Tables

Chapter 5. List of Figures

5.1 Introduction and Scope

Carbon Capture Utilization and Storage (CCUS) is the process of capturing CO_2 before it is emitted into the atmosphere, compressing and transporting the CO_2 to a geologic storage site, and injecting it into the site for storage [4]. The geologic storage site could be a deep saline formation, a depleted oil field, or an active oil field conducting CO_2 enhanced oil recovery (EOR). CCUS is based on practices and technologies developed in oil and gas exploration and production and natural gas storage, and is considered to be a potentially effective technology for the reduction of CO_2 emissions from stationary sources.

The 2018 increase in value of tax credits for CCUS in section 45Q of the US tax code provides a mechanism for CO_2 sources to monetize their emissions through tax credits [5]. A comprehensive evaluation of the feasibility of CO_2 storage projects for the purposes of securing 45Q tax credits requires detailed site characterization, geologic understanding, reservoir modeling, and monitoring design. However, during the early stages of planning, the information needed for in-depth project assessment may simply not be available to potential project developers, or the confidence in the available data may be low and/or the information may be broad or unrefined. Also, regulators faced with the review of detailed project technical reports may prefer simpler tools to bound the projected probabilistic performance of the reference design. Thus, there is a need for reliable (i.e., reasonably accurate) screening models to predict such performance metrics as pressure buildup and the spatial extent of CO_2 plume – while requiring only a limited amount of information that can be readily obtained from the literature and/or analog sites.

Motivated by these goals, a number of screening models have been presented in the literature for rapid performance assessment of CO_2 geologic sequestration projects. One set of studies builds upon Buckley-Leverett type frontal advance theory-based models of injection well pressure buildup and CO_2 plume displacement in simplified geometries [19, 3, 6, 16]. Another set of studies is based on vertical equilibrium type models [14, 9, 20]. A third approach is based on using full-physics based models to develop reduced-order models with physically relevant coefficients and/or dimensionless groups Each of these studies requires different levels of inputs about the system under consideration, and also exhibits varying levels of accuracy when compared to detailed simulation model results.

The objective of this study is to re-visit our earlier work [17, 16]. and provide an updated set of correlations that require even less information than the earlier screening models (and representative of the level of knowledge typical of project planning phases). The key idea is to use outputs from physics-based models to develop a simplified model of acceptable accuracy. First, we present a description of the reference model and inputs therein used to generate the synthetic data. Next, we present a brief overview of the key concepts used to develop the previous set of correlations. Then, the development and validation of the correlations for our screening model is described, along with its application to two example cases.

The paper ends with some concluding remarks regarding the applicability of the new screening model and its limitations.

5.2 Dataset Used for Analysis

In this study, the synthetic (simulation-generated) data used for developing the screening model are taken from an earlier work focusing on the development of simplified physics-based, statistical learning-based and reduced-order method based modeling approaches [12]. Specifically, this dataset has been used to develop correlations for pressure buildup [17] and spatial extent of CO_2 plume migration [18]. These correlations will be used as starting points for the present analysis.

The system being studied represents a single-well injecting supercritical CO_2 into a bounded 2-D radial-cylindrical formation initially filled with brine. The model domain consists of a porous and permeable heterogeneous reservoir, overlain by a low-permeability cap rock. The top of the cap rock, the bottom of the reservoir and the lateral boundary are all assumed to be no-flow boundaries. The simulations are executed in the numerical simulator Generalized Equation of state Model GEM® developed by the Computer Modeling Group. GEM is a robust, multidimensional and fully compositional reservoir simulator that is widely used to model the flow of multiphase and multicomponent fluids in the oil and gas industry, as well as equivalent problems in the context of geologic carbon storage.

The simulations were based on an experimental design that included a reference case, and multiple one-off sensitivity analyses carried out to perturb the most relevant parameters over a range. The selected discrete states for various parameters of interest are described below in **Table 5-8**. Note that the last parameter, which is the slope of the fractional flow curve corresponding to the CO_2-brine relative permeability relationship (i.e., a characterization of two-phase flow properties), is explained in further detail in the next section.

Table 5-8. MG-GEM Inputs for Experimental Design

Parameter	Experimental Design Values			Description
	Reference	Case A	Case B	
h	150	50	250	Reservoir thickness (m)
h_{CR}	150	100	200	Caprock thickness (m)
k	46	12	220	Avg reservoir permeability (mD)
$\sigma_{\ln k}$	3.56	2.45	4.67	Standard deviation of reservoir ln(k)
V_{DP}	0.55	0.35	0.75	Dykstra-Parsons coefficient
$k_{avg,CR}$	0.02	0.002	0.2	Average caprock permeability (mD)
k_V/k_{avg}	0.1	0.01	1	Anisotropy ratio
q	0.83	0.33	1.33	{Injection rate (million MT/yr)
L	10	5	7	System size (km)
ϕ	0.12	0.08	0.18	Reservoir porosity (fraction)
ϕ_{CR}	0.07	0.05	0.1	Caprock porosity (fraction)
l_k	Random	Increase from top	Decrease from top	Nature of permeability layering in the vertical direction
df_g/dS_g	1.74	4.74	6.24	Slope of CO_2 fractional flow curve (see explanation below)

5.3 Basic Concepts and Previous Work

5.3.1 Fractional Flow Curves

The concept of fractional flow is related to the relative permeability relationships between the gas (CO_2) and water (brine) phases during two-phase displacement processes, and is defined as follows:

$$f_g = \frac{\frac{k_{rg}}{\mu_g}}{\frac{k_{rg}}{\mu_g} + \frac{k_{rw}}{\mu_w}} \qquad \textbf{(Eqn. 1)}$$

where k_{rj} is the relative permeability of phase j, and μ_j its viscosity [8]. **Figure 5-21** shows a typical set of gas-water relative permeability curves, and the corresponding CO_2 (gas) fractional flow curve. Also shown therein is the tangent drawn to the CO_2 fractional flow curve. The slope of this line at the point of tangency df_g/dS_g gives the velocity of the Buckley-Leverett front (i.e., saturation-boundary between the injected CO_2 and in-situ brine), and its reciprocal is related to the average CO_2 saturation in the two-phase region [8].

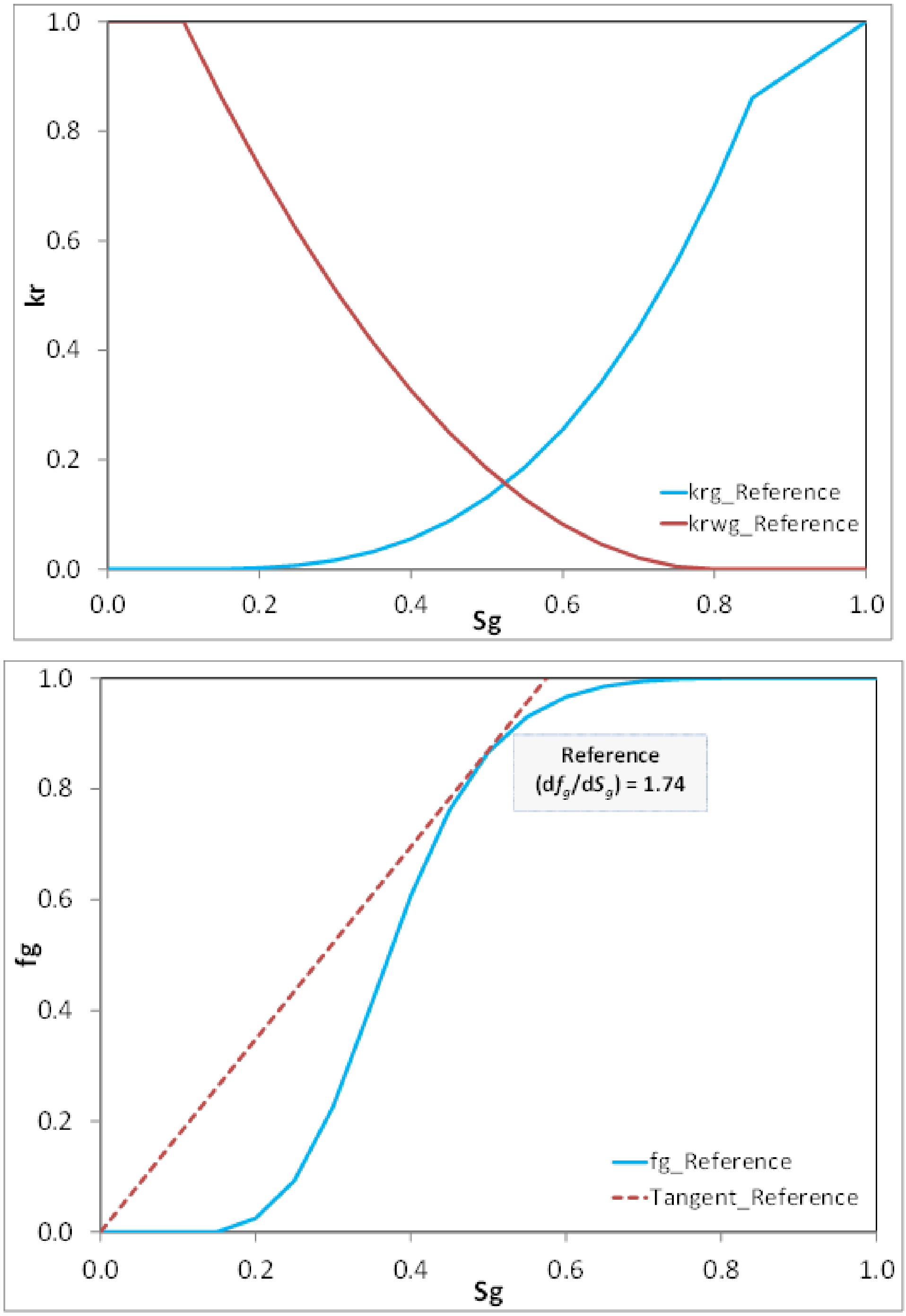


Figure 5-21. . Example relative permeability relationships (left) and fractional flow curve (right).

The slope of the CO_2 fractional flow curve is one of the hardest inputs to obtain for any given reservoir due to the inherent uncertainty in the relative permeability models. This parameter is generally approximated using laboratory experiments with core data. In the absence of core data to test with, literature-based *d*f*g*/*d*S*g* values can be useful assumptions to bound the associated uncertainty. For example, Bachu [1] and Bennion and Bachu [2] report relative permeability data using laboratory core experiments of samples from various formations in Canada that are candidates for CO_2 sequestration projects. They also developed useful generalizations for sandstone and carbonate formations respectively. In the absence of site-specific data, these databases can be used to construct the gas-water relative permeability curves, calculate the fractional flow curve, and obtain its slope. The use of data from analog sites/reservoirs is also another option for generating this information.

5.3.2 Injectivity Index

Injectivity index is a commonly used concept in petroleum reservoir engineering (similar to productivity index) for evaluating an injection well's capability for injecting fluids into a permeable formation. It is defined as the ratio of the injection rate q divided by the pressure difference between a reference formation pressure P_i and stabilized bottom-hole pressure P_f [21], i.e.,

$$J = \frac{q}{(P_f - P_i)} \qquad \textbf{(Eqn. 2)}$$

The utility of knowing the value of the injectivity index is that the stabilized pressure buildup corresponding to any target rate (or equivalently, the rate corresponding to a target pressure buildup) can be easily calculated without doing any detailed analytical or numerical modeling.

In the case of CO_2 injection into deep saline formations with open boundaries [11], or during CO_2 injection into depleted oil fields before the onset of boundary effects [13], it has been shown that a plot of the nominal injectivity index against time generally reaches a stabilized value that is proportional to the permeability-thickness product of the formation.

Ravi Ganesh and Mishra [17] used the concept of a dimensionless pressure group to describe this "pressure jump", i.e. the stabilized pressure increase corresponding to a constant rate of injection,

$$\mathrm{P_D} = \frac{2\pi kh}{\mu_w}\frac{(P_f - P_i)}{q} = \frac{2\pi kh}{\mu_w}\frac{1}{J} \qquad \textbf{(Eqn. 3)}$$

where kh is the reservoir permeability-thickness product, and μ_w is viscosity of the native brine (displaced fluid). From P_D, injectivity index J can be readily calculated knowing permeability-thickness product kh and brine viscosity μ_w. Ravi Ganesh and Mishra [17] showed that P_D could be expressed as a quadratic model for a wide-range of formation properties as well as two-phase relative permeability relationships.

$$P_D = 10.3 + 0.59\frac{df_g}{dS_g} + 3.41V_{DP} + 1.23\frac{df_g}{dS_g}V_{DP} - 0.342\left(\frac{df_g}{dS_g}\right)^2 - 8.89(V_{DP})^2 \qquad \textbf{(Eqn. 4)}$$

Here, df_g/dS_g is the slope of the CO_2 fractional flow curve discussed earlier, and V_{DP} is the Dykastra-Parsons coefficient commonly used in the oil industry for characterizing log-normal permeability distributions [8]:

$$V_{DP} = \frac{(k_{50} - k_{84.1})}{k_{50}} \quad \textbf{(Eqn. 5)}$$

where k_{50} is the median permeability, and $k_{84.1}$ is the permeability value that is one standard deviation away from the median. Values of V_{DP} range between from 0 (perfectly homogenous) to 1 (completely heterogeneous). V_{DP} can also be shown to be equal to (1-exp(-σ_{lnk})), where σ_{lnk} is the standard deviation of the natural log of permeability values [13] as noted earlier in **Table 5-8.** Note that at a project screening level, the information needed to calculate V_{DP} (e.g., using values of individual permeability samples fitted to a log-normal distribution) may not always be available.

5.3.3 Total Storage Efficiency

As noted by several researchers [8, 14] for the case of displacement of a dense, viscous fluid (e.g., brine) by a less dense and less viscous fluid (e.g., CO_2), the migration of the injected fluid is controlled by mainly the effects of buoyancy (gravity), mobility/viscosity contrast between the fluids, and reservoir heterogeneity. Based on the schematic in **Figure 5-22** taken from Ravi Ganesh and Mishra [18], we note that within the footprint of the CO_2 plume only a fraction of the total pore volume in the reservoir is contacted by CO_2. The ratio of the volume swept (contacted) by CO_2 to the total volume within the footprint of the CO_2 plume is defined as the volumetric sweep efficiency (E_v). Within this swept volume, the efficiency of displacement of the native brine by the CO_2 is given by the displacement efficiency. As the initial gas saturation in the reservoir is zero, the average CO_2 saturation behind the front, $S_{g,av}$, gives this displacement efficiency. The total storage efficiency, defined as the product of E_v and $S_{g,av}$, thus signifies the efficiency of CO_2-brine displacement process or the ability to effectively sequester CO_2 in that reservoir (E_s). The lower the value of total storage efficiency, the greater is the extent of the plume from the injection well.

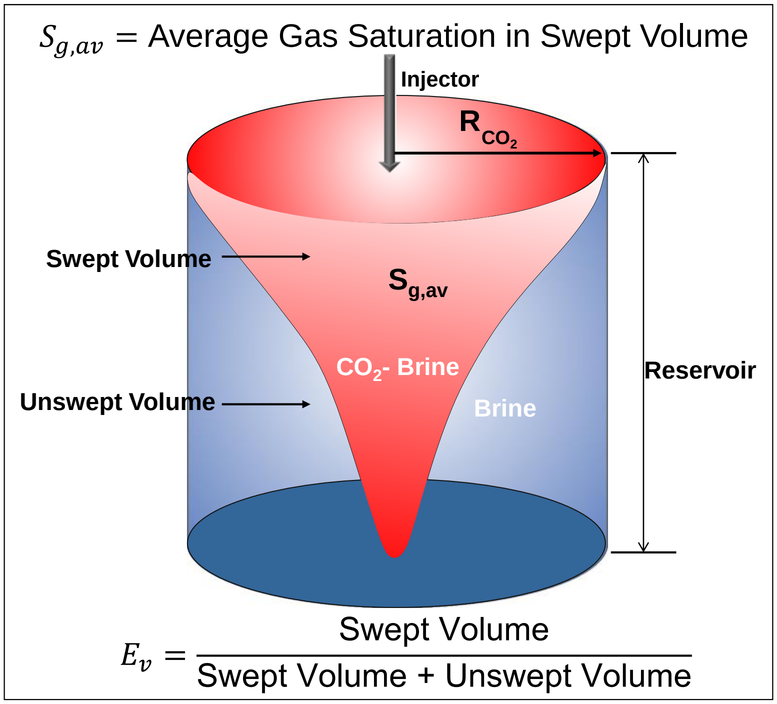


Figure 5-22. System schematic showing graphical definitions of plume extent, volumetric sweep and displacement efficiency for CO_2- brine displacement [18].

The total storage efficiency is related to the maximum radial extent of the CO_2 plume at the end of injection, R_{CO2}, as given in Eqn. 6 below:

$$R_{CO2}^2 = \frac{Q}{\pi\phi h S_{g,av} E_v} = \frac{Q}{\pi\phi h E_s} \quad \textbf{(Eqn. 6)}$$

where $Q = qtB_g(1-x_f)$ is the cumulative reservoir volumes of CO_2 injected with q being the surface mass injection rate for time t, B_g the CO_2 formation volume factor (i.e., ratio of reservoir volumes to surface volumes), and x_f is the mass fraction of CO_2 dissolved in brine. Using simulated data for a broad range of conditions, Ravi Ganesh and Mishra [18] developed a quadratic model for the total storage efficiency, E_s, as follows:

$$E_S = 30.7 + 0.435\frac{df_g}{dS_g} + 29.24L_C - 22.02V_{DP} - 11.2N_g + 4.59\frac{df_g}{dS_g}V_{DP} - 25.21L_CV_{DP} - 0.692\left(\frac{df_g}{dS_g}\right)^2 + 6.11{N_g}^2 \quad \textbf{(Eqn. 7)}$$

As noted earlier, df_g/dS_g is the slope of the CO_2 fractional flow curve, and V_{DP} is the Dykastra-Parsons coefficient. The other two dimensionless parameters, gravity number N_g, and Lorenz coefficient, L_c, are described below.

The gravity number N_g signifies when gravity effects are more pronounced compared to viscous flow effects and vice-versa [8]. This quantity plays an important role to characterize different stages of geologic sequestration due to the density difference between the more buoyant supercritical CO_2 and the heavier native brine [14]. Although the definition for gravity number in surveyed literature varies from source to source [15], Ravi Ganesh and Mishra [18] used the following definition of the gravity number while accounting for reservoir permeability anisotropy in developing their correlation:

$$N_g = \frac{(\Delta\rho g h) k h (\frac{h}{L})}{q \mu_g \left(\frac{k_V}{k}\right)} \qquad \textbf{(Eqn. 8)}$$

where $\Delta\rho = (\rho_w - \rho_g)$ is the difference in brine and CO_2 densities, g is acceleration due to gravity, μ_g is the CO_2 viscosity, (k_V/k) is the vertical to horizontal permeability anisotropy ratio, L is system dimension, and other terms are as defined earlier. Note that the gravity number can readily computed if fluid properties can be estimated from the corresponding pressure and temperature values, and typical reservoir properties are known.

The Lorenz coefficient L_C yields another dimensionless characterization of the reservoir heterogeneity. It is defined as [8]:

$$\mathrm{L_C} = 2\left\{\int_0^1 \mathrm{F_n dC_n} - \frac{1}{2}\right\} \qquad \textbf{(Eqn. 9)}$$

where F_n is the cumulative flow capacity (i.e., fraction of the total permeability-thickness product up to a given layer), C_n is the cumulative storage capacity (i.e., fraction of the total porosity-thickness product upto a given layer), and L_c is a common measure of heterogeneity that is given by the area between the F-C curve and a 45^0 line (homogeneous F-C curve) and normalized by 0.5. As in the case of V_{DP}, L_C ranges from 0 to 1 with a value *of* zero indicating homogeneous reservoirs and values closer to one indicating extremely heterogeneous ones. It should also be noted that at a project screening level, the information needed to calculate L_c (i.e., detailed permeability and porosity values on a layer by layer basis) may not always be available.

5.3.4 New Correlations

As mentioned in the Introduction, the primary motivation for developing new correlations for J and E_S is to eliminate the dependence on parameter groups such as V_{DP} and L_c which require a detailed knowledge of reservoir porosity and permeability distributions. The development and validation of these new correlations is presented below.

5.3.5 Injectivity Index

As described previously, Ravi Ganesh and Mishra [17] have shown that the dimensionless pressure buildup following CO_2 injection can be correlated as a function of the slope of the CO_2 fractional flow curve df_g/dS_g, and the Dykstra-

Parsons coefficient, V_{DP}. In **Figure 5-23** the same data is presented in a dimensional form with the injectivity index J (in MT/yr/psi) plotted against the permeability-thickness product kh (in mD-ft). The figure suggests that the trend in the data can be captured via a simple linear relationship of the form:

$$J = \alpha\,(kh) \qquad \textbf{(Eqn. 10)}$$

where the constant of proportionality α is primarily dependent on the relative permeability relationship used for the simulations, with a secondary dependence on permeability heterogeneity.

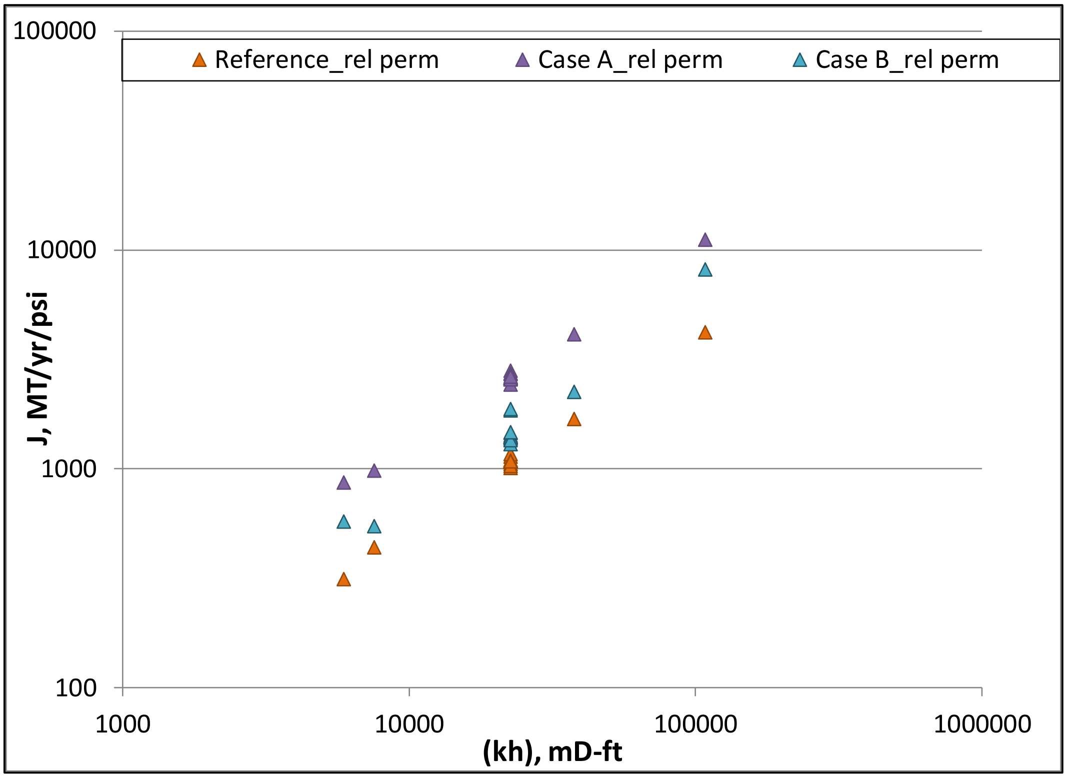


Figure 5-23. Simulated values of injectivity index correlated to permeability-thickness product for different relative permeability curves [17].

We postulate that the slope of the fractional flow curve df_g/dS_g is an appropriate parameterization of the complex relative permeability relationship. As a first-order approximation, using the data from **Figure 5-23** above, the constant of proportionality in Eqn. 10 can be expressed as:

$$\alpha = 0.033 * \exp(0.2*df_g/dS_g) \qquad \textbf{(Eqn. 11)}$$

where df_g/dS_g is used as a single parameter to capture the essence of the two-phase relative permeability curves, and the influences of other potential factors such as permeability heterogeneity are ignored. Thus, the new correlation for injectivity index can be stated as:

$$J = 0.033 * \exp(0.2*df_g/dS_g) * kh \qquad \textbf{(Eqn. 12)}$$

To validate this simple correlation, it has been used to predict the injectivity index for all cases in the dataset described by Ravi Ganesh and Mishra [17]. A cross-plot of the simulated and predicted *J* values, shown in **Figure 5-24**, indicates excellent agreement with a coefficient of determination R^2=0.98.

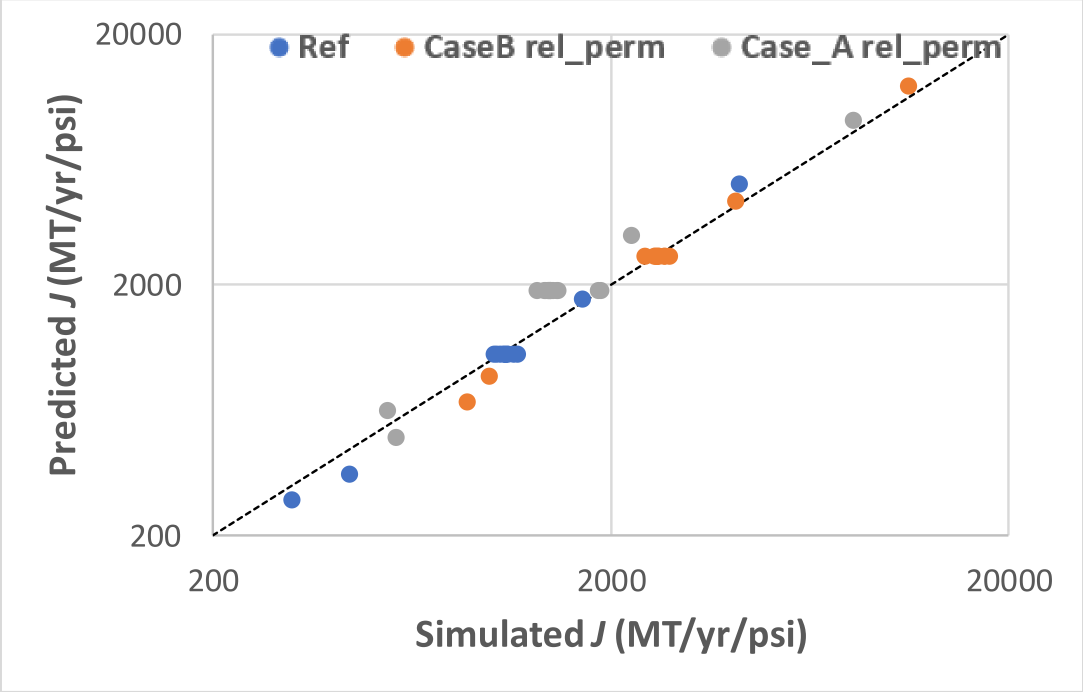


Figure 5-24. Comparison between simulated and predicted injectivity index values, showing excellent agreement, where the colors refer to the three sets of relative permeability relationships.

In summary, we have developed a simple, but accurate, correlation for injectivity index *J* in terms of permeability-thickness product *kh* and slope of the CO_2 fractional-flow curve df_g/dS_g. Knowing *J*, the injection well pressure buildup for a given injection rate (or vice versa) can be readily calculated using Eqn. 2. Note that this correlation was developed over the range *kh* (mD-ft) = [5800, 110000], and df_g/dS_g = [1.7, 6.3]. Caution should be exercised when extrapolating the results beyond these ranges.

5.3.6 Total Storage Efficiency

As noted earlier, Ravi Ganesh and Mishra [18] have shown that the total storage efficiency can be correlated as a function of the slope of the CO_2 fractional flow curve df_g/dS_g, the Dykstra-Parsons coefficient V_{DP}, the Lorenz coefficient L_c, and the gravity number N_G. In **Figure 25** the same data is presented with the total storage efficiency E_s plotted against gravity number N_G, which suggests a general linear trend of the form

$$E_S = \beta + \gamma\ (\ln(N_G)) \qquad \textbf{(Eqn. 13)}$$

for each distinct relative permeability relationship, albeit with some degree of scatter related to other variables that characterize reservoir heterogeneity.

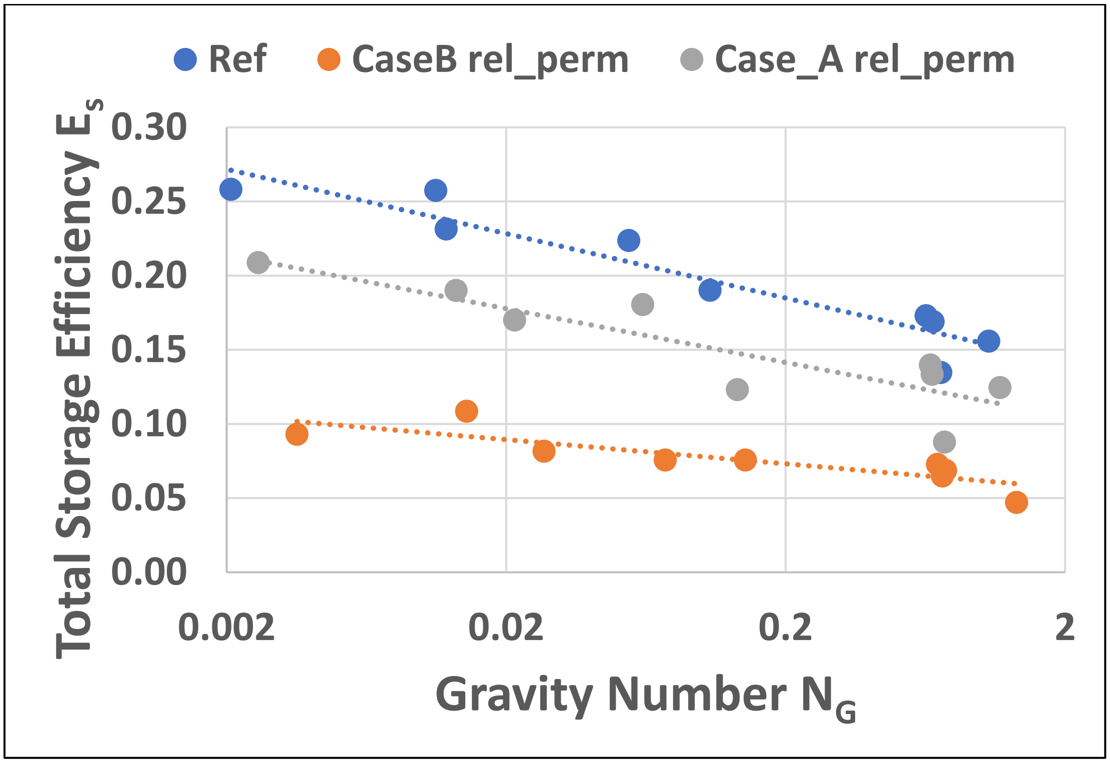


Figure 25-5. Simulated values of total storage efficiency correlated to gravity number for different relative permeability curves [18].

We hypothesize that as a first-order approximation, the trend lines drawn in Figure 25 can be captured using a single parameter df_g/dS_g while subsuming the influence of parameter groups such as V_{DP} and L_C. Using the data from Figure 25, the two constants in Eqn. 13 above, β and γ, can be expressed as simple functions of df_g/dS_g as follows:

$$\beta = 1.56\text{E-}1 - 3.90\text{E-}4\ (df_g/dS_g)^3 \qquad \textbf{(Eqn. 14)}$$

$$\gamma = -1.93\text{E-}2 - 8.07\text{E-}6\ (df_g/dS_g)^4 \qquad \textbf{(Eqn. 15)}$$

To validate the correlations presented in Eqn. 13 through Eqn. 15, they have been used to predict the total storage efficiency for all cases in the dataset described earlier. A cross-plot of the simulated and predicted E_S values, shown in

Figure 5-26, indicates very good agreement with a coefficient of determination R^2=0.95.

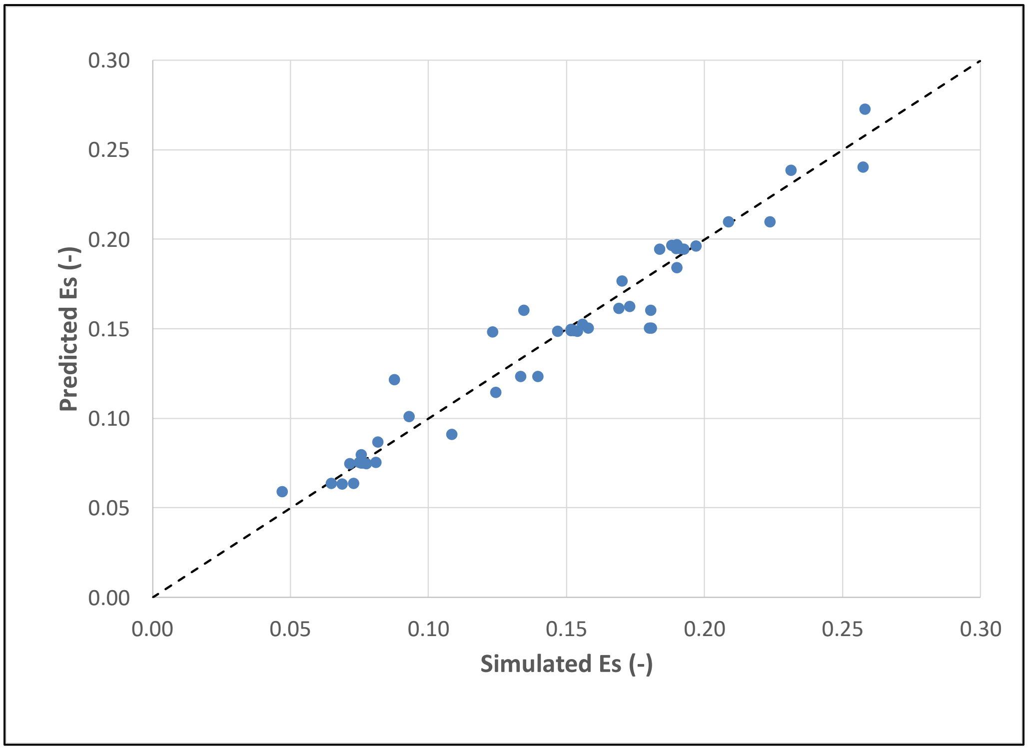


Figure 5-26. Comparison between simulated and predicted total storage efficiency values, showing excellent agreement.

In summary, we have developed a simple, but accurate, correlation for total storage efficiency E_s in terms of a modified gravity number N_g and the slope of the CO_2 fractional-flow curve df_g/dS_g. Knowing E_s, the CO_2 plume radius can be readily calculated using Eqn. 6. Note that this correlation was developed over the range N_G = [0.002, 2], and df_g/dS_g = [1.7, 6.3]. Caution should be exercised when extrapolating the results beyond these ranges.

5.4 Example Applications

The application of the new correlations involves two steps: (a) calculate injectivity index J, from Eqn. 12, to determine the pressure buildup and final quasi-stable pressure, and (b) calculate gravity number N_g, from Eqn. 8, and total storage efficiency E_s, from Eqn. 13 through Eqn. 15, to determine radius of CO_2 plume R_{CO2}. This requires calculating a number of fluid properties for brine and CO_2 from correlations. For example, density of CO_2 ρ_g; viscosity of CO_2 μ_g; and formation volume factor of CO_2 B_g can be estimated from NIST tables [7]. Mass fraction of CO_2 in brine x_f, can be assumed to vary between 0.05 and 0.08 (based on simulations and literature values). Density of formation brine ρ_w, and viscosity of formation brine μ_w, can be estimated from the correlations presented in McCain [10].

This workflow has been used to calculate P_f and R_{CO2} for two example cases described below. The first (Example #1) is a STOMP simulation for the “warm-shallow” case described in Oruganti and Mishra [16]. The second (Example #2) is a CMG-GEM simulation used as the blind validation #2 case in Ravi Ganesh and Mishra [18]. The relevant parameters for the two cases are shown below in **Table 5-9**. For instance, the calculated intermediate parameters used for Example #2 are $\rho_g = 845.9$ kg/m^3; $\mu_g = 0.085$ cP; $B_g = 0.3909$ bbl/Mcf; $x_f = 0.08$; and $\rho_w = 972.9$ kg/m^3.

Table 5-9. Input values for the two example cases.

Parameter	Description	Units	Example #1	Example #2
P_i	Initial pressure	psi	1615	1839
T_i	Initial temperature	deg F	153	101
Sal	Brine salinity	molal	.15	.15
L	Reservoir lateral extent	m	57950	7500
h	Reservoir thickness	m	100	148.8
ϕ	Reservoir average porosity	--	.1	.1451
k	Reservoir average horizontal permeability	mD	100	58.25
k_v/k	Reservoir anisotropy ratio	--	1	.2482
df_g/dS_g	Slope of CO_2 fractional flow curve	--	3.49	1.74
q	CO_2 mass injection rate	MT/yr	1.5e6	1.33e6
t	Injection time	yr	5	30

The final results for these two cases, along with the corresponding simulated values, are presented below in **Table 5-10**, indicating excellent agreement. This demonstrates the practical utility of the new correlations for screening-level calculations with only limited information about the reservoir.

Table 5-10. Results for the two example cases.

	Example #1	**Example #2**
Simulated R_{CO2} from numerical model, m	1250	1699
Calculated R_{CO2} from Eq. 13-15, and Eq. 6 (this study), m	1287 (error 3.0%)	1743 (error 2.6%)
Simulated P_f from numerical model, psi	2420	2881
Calculated P_f from Eq. 12 and Eq. 2 (this study), psi	2298 (error 5.0%)	2830 (error 1.8%)

5.5 Concluding Remarks

The objective of this short paper is to revisit the correlations presented in Ravi Ganesh and Mishra [17] for injection-well pressure buildup and Ganesh and Mishra [18] for CO_2 plume migration to develop a new screening model. To this end, we have developed a new correlation for injectivty index J in terms of the slope of the CO_2 fractional flow cure, df_g/dS_g. A second new correlation has been developed for total storage efficiency within the footprint of plume E_s as a function of gravity number N_g and df_g/dS_g. Using these two correlations, as well as a knowledge of some basic reservoir characteristics (e.g., Table 5-9) and estimates of fluid properties from standard correlations, the injection-well pressure buildup and CO_2 plume extent in the formation can be readily estimated. The new correlations reproduce the results of the underlying simulations quite well, and also provide good agreement with independent calculations for the two example problems.

The primary limitations of this screening model are essentially those of the simulation model which created the original data [12], i.e.: (a) areal homogeneity, (2) radial geometry, and (3) single well configuration. Also, the pressure buildup is captured only at the injection well, and its radial extent (i.e., pressure plume) is not quantified. There are also several conditions in the original data (i.e., vertical heterogeneity, caprock-reservoir system) which are ignored in the present model. However, as shown in Figure 5-24 and Figure 5-26, these simplifications do not appear to materially affect the accuracy or robustness of the new correlations.

In conclusion, the primary contribution of this work is a new screening model for predicting injection-well pressure buildup and CO_2 plume migration of CO_2 geologic sequestration projects. The model requires only limited information and is quite accurate (when compared to detailed simulation results). We believe the model can be a potentially useful tool for project developers during the early days of project planning (e.g., for 45Q related projects), and also for regulators looking for a simple check against detailed numerical models.

5.6 References

1. Bachu, S. 2013. Drainage and Imbibition CO_2/Brine Relative Permeability Curves at in Situ Conditions for Sandstone Formations in Western Canada. Energy Procedia, Volume 37, pp. 4428-4436.
2. Bennion DB and Bachu S. 2010. Drainage and imbibition relative permeability curves at reservoir conditions for carbonate formations. SPE Paper 134028, SPE Annual Technical Conference and Exhibition, 19-22 September 2010, Florence, Italy.
3. Benson, S. M. 2003. Pressure buildup at CO_2 injection wells, Second Annual Conference on Carbon Sequestration, Alexandria, VA, May 5–8.
4. Benson SM and Friedman SJ, Carbon Dioxide Capture, Utilization and Storage: An Important Part of a Response to Climate Change. National Academy of Engineering, Washington, DC, pp. 42–50 (2014).
5. Bomgardner, 2020. https://cen.acs.org/environment/greenhouse-gases/45Q-tax-credit-s-luring/98/i8.
6. Burton, M., Kumar, N., and Bryant, S. 2008. Time-Dependent Injectivity during CO_2 Storage in Aquifers. Paper SPE-11937 presented at the 2008 SPE/DOE Improved Oil Recovery Symposium, Tulsa, Oklahoma, 19–23 April.
7. Jarrell, P.M., Fox, C.E., Stein, M.H., Webb, S.L. 2002. Practical Aspects of CO_2 Flooding. SPE Monograph Series. Society of Petroleum Engineers, v. 22, 220 p.
8. Lake, L. W., Johns, R. T., Rossen, W. R., & Pope, G. 2015. Fundamentals of enhanced oil recovery, Richardson, TX, Society of Petroleum Engineers, 496 p.
9. Mathias, S., Hardisty, P., Trudell, M., and Zimmerman, R. 2008. Approximate Solutions for Pressure Buildup During CO2 Injection in Brine Aquifers. Transport in Porous Media, no. 2. pp. 265–284
10. McCain, W. J.. 1989. "The Properties of Petroleum Fluids," Tulsa, OK, PennWell Books.
11. Mishra, S., M. Kelley, E. Zeller, N. Slee, N. Gupta, I. Bhattacharya and M. Hammond, 2013. Maximizing the value of pressure monitoring data from CO_2 sequestration projects. Energy Procedia 37 (2013) 4155-4165.
12. Mishra, S., Ravi Ganesh, P., Schuetter, J., He, J., Jin, Z., and Durlofsky, L.J. 2015. Developing and validating simplified predictive models for CO_2 geologic sequestration, SPE-175097, ATCE, Sept. 28-30.
13. Mishra, S. Ravi Ganesh, P., Schuetter, J. 2017. Developing and Validating Simplified Predictive Models for CO_2 Geologic Sequestration. Energy Procedia, Vol. 114, pp. 3456-3464.

14. Nordbotten, J. M., and Celia, M. A. 2006. Similarity solutions for fluid injection into confined aquifers, J. Fluid Mech., 561, pp. 307–327.

15. Novakovic, D. 2002. Numerical reservoir characterization using dimensionless scale numbers with application in upscaling. LSU Doctoral Dissertations.
443. https://digitalcommons.lsu.edu/gradschool_dissertations/443

16. Oruganti Y, and Mishra S. 2013. An improved simplified analytical model for CO_2 plume movement and pressure buildup in deep saline formations. International Journal of Greenhouse Gas Control, Vol. 14, p. 49-59; DOI: 10.1016/j.ijggc.2012.12.024.

17. Ravi Ganesh, P. and Mishra, S. 2014. Reduced physics models of CO2 injectivity. Energy Procedia, Vol. 63, pp. 3116-3125.

18. Ravi Ganesh, P. and Mishra, S. 2015. Simplified physics model of CO_2 plume extent in stratified aquifer-caprock systems. Greenhouse Gases: Science and Technology, 6(1), pp.70-82

19. Saripalli K.P. and McGrail B.P. 2002. Modeling deep-well injection of CO_2 for geological sequestration. Energy Conversion and Management, Vol. 43, p. 185–198.

20. Swickrath, M., S. Mishra, and P. Ravi Ganesh, 2015. A critical evaluation of sharp interface models for CO2-brine displacement in saline aquifers, Groundwater, DOI: 10.1111/gwat.12366.

21. Terry, R.E and Rogers, J.B. 2014. Applied Petroleum reservoir Engineering 3rd edition. Upper Saddle River, N.J., Prentice Hall, 528 p.

Chapter 6.

Effect of Injection Well Inclination on Injectivity and Pressure Response

Priya Ravi Ganesh and Srikanta Mishra

Chapter 6. Table of Contents

Chapter 6. List of Tables

Chapter 6. List of Figures

6.1 Introduction

Vertical wells are the conventional injection well configuration employed for CO_2 injection in Carbon Capture Utilization and Sequestration (CCUS) projects. The primary objective of this study is to evaluate the performance of horizontal well configurations as a potential alternative in suitable systems. The system of interest is a synthetic numerical model representing a typical depleted reservoir in which CO_2 injection for geologic storage is simulated after primary recovery. Injectivity-index analysis approaches are applied to determine the rapid performance assessment of injection well configurations in reservoirs of varying permeabilities.

The report summarizes the model setup and the scenarios considered for the numerical analyses. This is followed by a discussion on the pressure response evaluation at potential monitoring well locations of interest and injectivity analyses using the simulated data.

6.2 Model Description

6.2.1 Geometry

The numerical model was based on 2-D radial geometry representing a typical closed oil reservoir that is ~400 m wide and ~91 m thick with a well at the center. This well acted as the producer during the primary production period and was converted into an injector during the CO_2 injection period. The medium permeability reservoir zone occupies the middle ~40 m of the model domain, overlain by 24 m of very low-permeability cap rock and underlain by a 31 m low-permeability water column. Figure 6-27 shows the model geometry and cross-sectional view of the porosity and permeability distributions in the different zones.

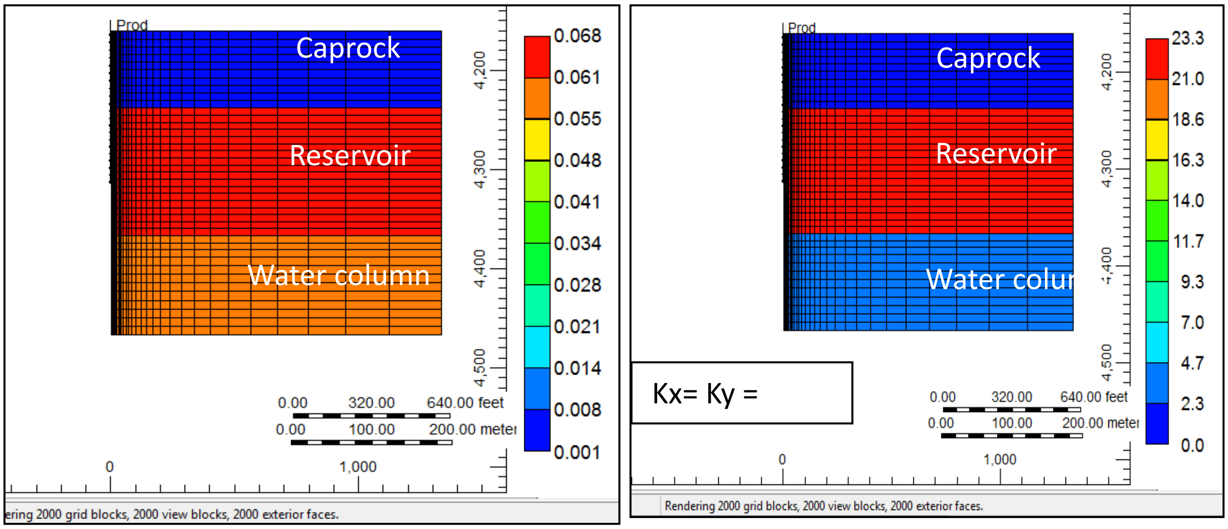


Figure 6-27. Cross-sectional view of the synthetic radial model showing the porosity (left) and permeability (right) distributions in the different zones. Grids are logarithmically increasing in size as we move away from the producing/ injecting well at the center of the model.

6.2.2 Rock and Fluid Properties

Table 6-11 summarizes the key petrophysical properties of the caprock, reservoir and underlying water column zones. The porosity of the 40 m reservoir zone was 7%. The porosities of the overlying caprock and underlying water column were 0.1% and 6% respectively, while their permeabilities were defined to be 1E-4 mD and 2 mD, respectively. Three reservoir permeability scenarios were evaluated at 8, 16 and 23 mD respectively.

Table 6-11. Petrophysical properties of the different zones in the modeled system of interest.

Petrophysical Property	Overlying Caprock zone	Middle Reservoir zone	Underlying water column zone
Thickness, m	24	40	31
Porosity, %	0.1	7	6
Permeability, mD	1E-4	8; 16; 23	2

The various hydrocarbon components were grouped into six pseudo-components: F1, F3, F4, F5, F6, and F7, along with one pseudo-component F2 for pure CO_2. Fluid property characterization was performed using these pseudo-components in CMG Winprop® to match the laboratory PVT data of a sample field chosen from Battelle's projects [1]. Table 6-12 lists the composition and molecular weights of the pseudocomponents in the in-situ hydrocarbon fluid phase modeled.

Table 6-12. Molecular weights of the pseudocomponents in the fluid modeled.

Component	Mol. Weight	Initial Mole Fraction
F1	16.116	0.409
F2 (CO_2)	44.010	0.001
F3	44.207	0.202
F4	95.459	0.174
F5	179.68	0.147
F6	297.181	0.028
F7	530.093	0.039

6.2.3 Simulation Scenarios and Metrics

The model was initialized at a pressure of ~20,000 kPa (~2900 psi) at a reference depth of ~1341 m (~4400 ft) with no initial gas cap present. Primary recovery period of 10 years was simulated during which 0.25 million standard cubic metres (1.6 MMSTB) or approximately 50% of the original oil in place was produced. Thereafter, CO_2 was injected at a constant rate of 500 metric tons per day using a gradual ramp-up schedule, i.e., (1) one month injection and two week shut-in, (2) two month injection and three week shut in, (3) three month injection and four week shut in, and (4)-(5) two additional periods of six month injection and four week shut in.

For each of the reservoir permeability scenarios, the performance of different injection well configurations were evaluated. The reference injection well configuration was a vertical well perforated through the reservoir zone with horizontal well configurations of different lateral lengths (reported in terms of half-lengths) tested for injectivity performance comparison. Table 6-13 shows the injection well configurations considered for each of the three reservoir permeability scenarios that results in 12 modeled scenarios of interest:

Table 6-13. Injection well configurations modeled for each of the three reservoir permeability scenarios in the synthetic aquifer model

Injection well configuration	Configuration description
Vertical well default reference	Vertical well perforated through
Horizontal well 1	240 ft half-length horizontal well through middle of reservoir zone
Horizontal well 2	400 ft half-length horizontal well through middle of reservoir zone
Horizontal well 3	670 ft half-length horizontal well through middle of reservoir zone

Figure 6-28 shows the injected CO_2 rate, the injection well bottomhole pressure and average pressure responses in the different reservoir permeability models with the reference vertical injection well configuration. The average reservoir pressure at the end of depletion and prior to start of the CO_2 injection phase is 5447 kPa (790 psi) which is seen to increase because of CO_2 injection into the closed system. These primary variables are used to evaluate the CO_2 injectivity index for the horizontal well performance assessment as explained in the subsequent section.

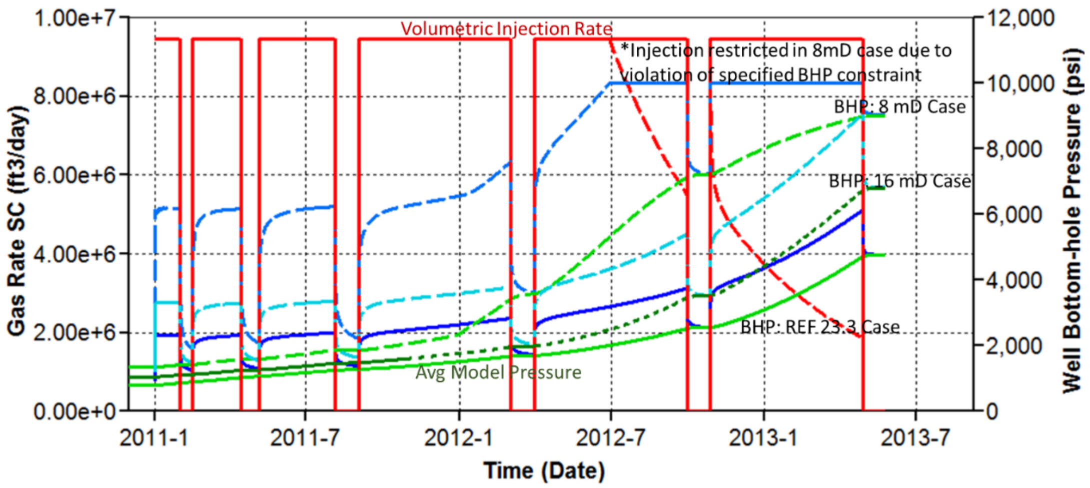


Figure 6-28. Injected CO_2 rate (red), the injection well bottomhole pressure (dashed curves) and average pressure responses (solid curves) in the different reservoir permeability models with the reference vertical injection well configuration

6.3 Results Analyses

6.3.1 Injectivity Analysis

Injectivity index is a simple and commonly used concept in petroleum reservoir engineering to evaluate the capability of a well to inject fluids into a porous and permeable formation [2]. It is defined as the ratio of the injection rate divided by the pressure difference between formation pressure and bottom-hole pressure:

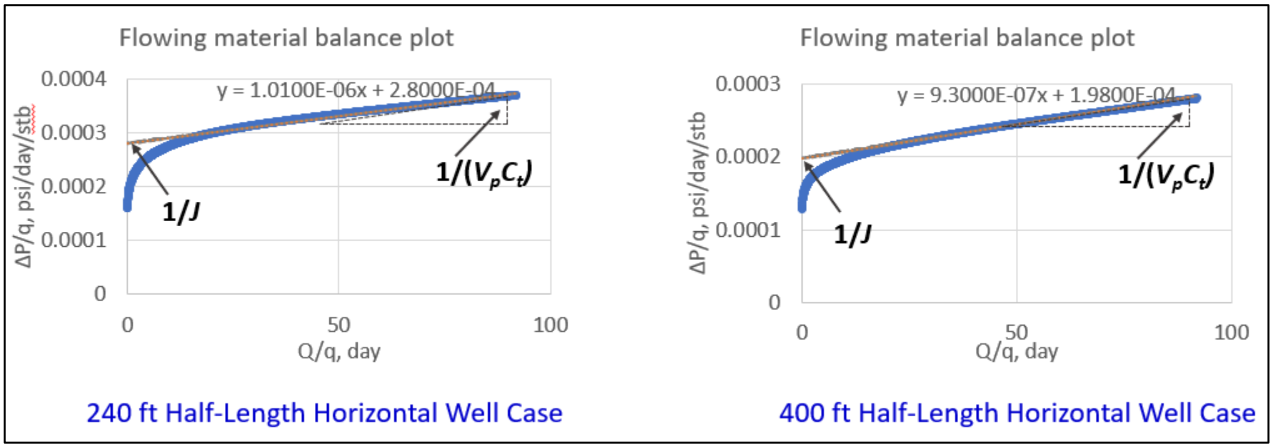
$$J = \frac{q}{(P_f - P_{BH})} \qquad \textbf{(Eqn. 1)}$$

where J is injectivity index, q is injection rate, P_f is reference average formation pressure and P_{BH} is bottom-hole pressure. For the pseudo-steady-state period during CO_2 injection following the initial transient period, the injectivity index can be calculated using the equation 2 as P_f is unknown.

$$\frac{P_i - P_{wf}}{q_{sc}} = \frac{P_i - \bar{P}}{q_{sc}} + \frac{\bar{P} - P_{wf}}{q_{sc}} = \left(\frac{Q}{q_{sc}}\right)\left(\frac{1}{V_p c_t}\right) + \frac{1}{J} \qquad \textbf{(Eqn. 2)}$$

where Q is the cumulative CO_2 injected, V_p is the pore volume and c_t is the total fluid compressibility.

This suggests that when injection well pressure build-up normalized by the injection rate is plotted against the ratio of cumulative injection to injection rate (i.e., material balance time), it should yield a straight line with slope inversely proportional to the pore volume times compressibility, and intercept equal to the reciprocal of the stable injectivity index. This is also referred to as a flowing material balance plot [3]. Figure 6-29 shows example calculations of the injectivity index from the flowing material balance plot for the injection period 3 in the 240 ft and 400 ft half-length horizontal well cases, where the injectivity index is calculated from the intercepts as 387 MT/yr/psi and 548 MT/yr/psi respectively.



240 ft Half-Length Horizontal Well Case

400 ft Half-Length Horizontal Well Case

Figure 6-29. Calculation of injectivity index, J using flowing material balance plot.

Figure 6-30 presents the calculated average injectivity indices for the 12 modeled scenarios of interest.

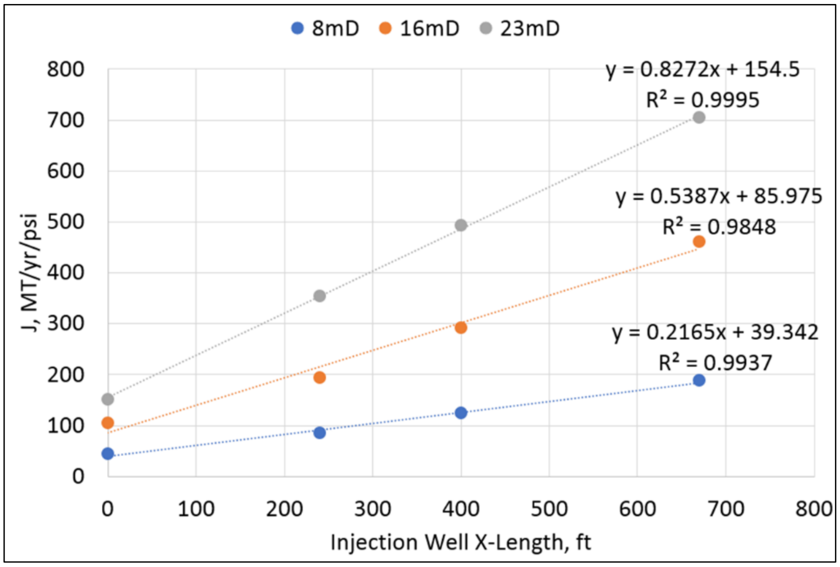


Figure 6-30. Calculated average injectivity index ratio, *J* for the modeled scenarios of interest.

The injectivity index for the reference vertical well configuration is the intercept value with the improved performance from horizontal wells demonstrated by the better injectivity values in the Figure 6-30. The injectivity index for the horizontal wells is observed to be directly proportional to the half-length of the lateral. Figure 6-31 normalizes the injectivity indices by representing a ratio of the horizontal well injectivity with respect to the reference vertical well at the corresponding reservoir permeability.

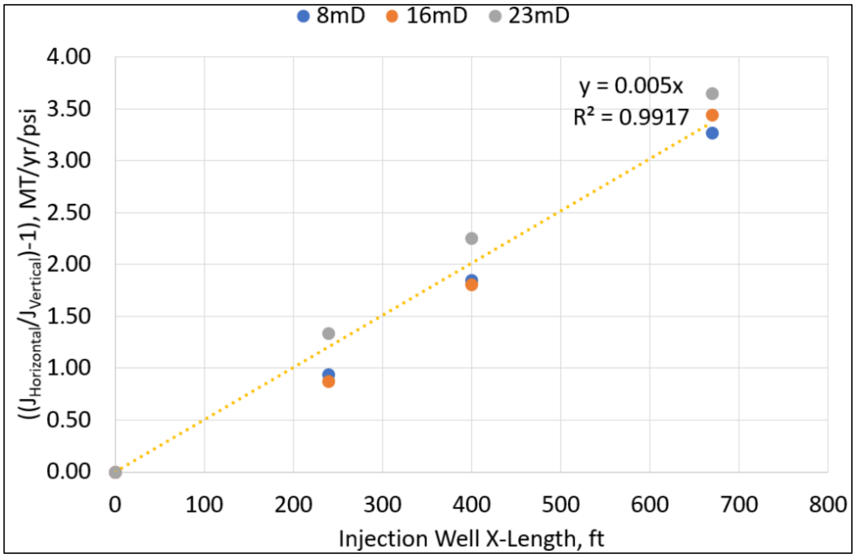


Figure 6-31. Calculated average injectivity index ratio, *J_{ratio}* for the modeled scenarios of interest.

The injectivity index ratio for a particular horizontal well is a function of the half-length of the lateral being considered and is given by a simple empirical correlation as:

$$J_{ratio}, \frac{\frac{MT}{yr}}{psi} = \frac{J_{horizontal}}{J_{vertical}} = 1 + \ 0.005.\,(Horizontal\ Injection\ well\ halflength, ft)$$

(Eqn, 3)

The variation in data (for the 23 mD reservoir permeability scenarios) around the correlation in Figure 6-31 reflects the uncertainty in the injectivity index due to the averaging process that was implemented. However, this simple correlation provides an extremely useful first pass estimate of horizontal well performance in a given reservoir that can be used for an evaluation of suitable well configurations preceding detailed numerical and design implementation in a geologic CO_2 sequestration project.

6.3.2 Pressure Response at Potential Monitoring Well Locations

While the pressure buildup per unit CO2 injected is lower for horizontal wells as reflected by the higher injectivity indices, pressure response at other locations in the system is also important to monitor the impact of injecting in a horizontal configuration in comparison to conventional vertical wells. The pressure response observed at two potential monitoring well locations of interest is shown in Figure 32. The two monitoring away from the end of the lateral.

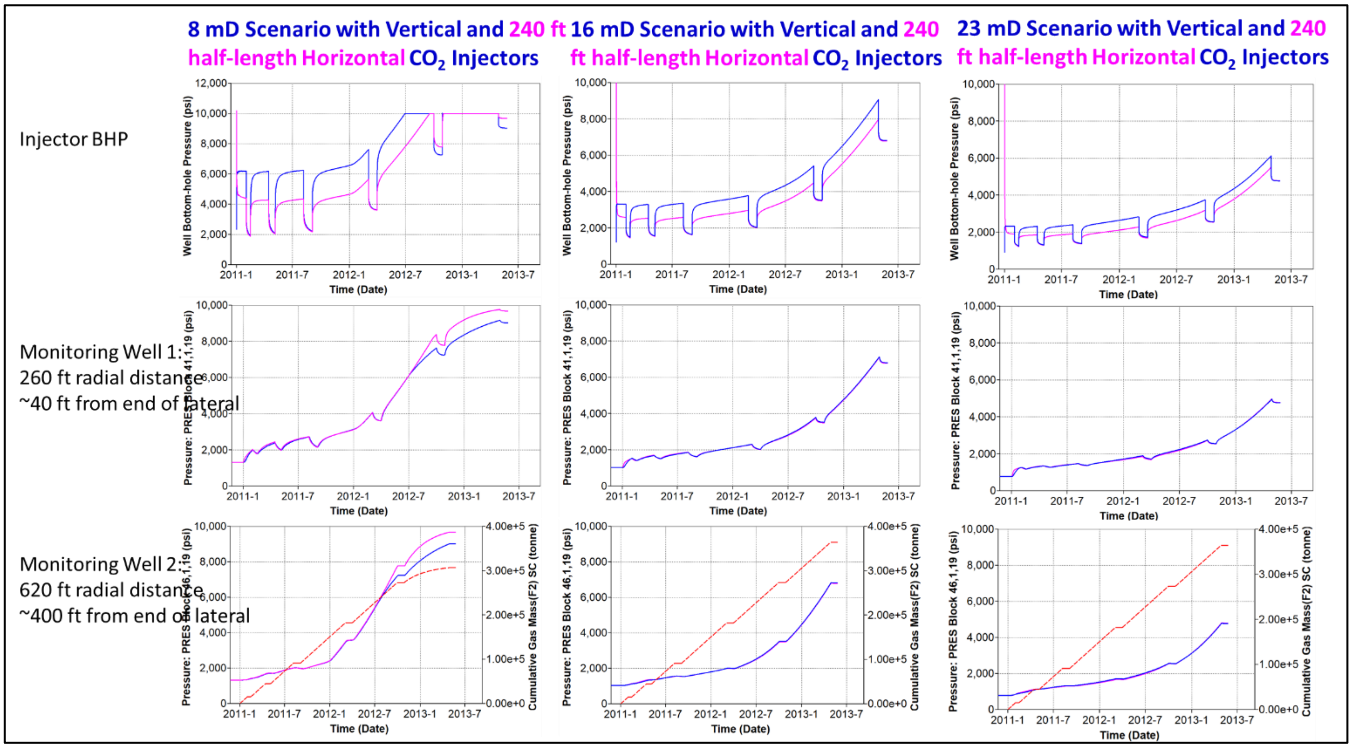


Figure 32-6. Bottomhole pressure responses at injection well and two potential monitoring well locations of interest just outside the end of and 400 ft from the lateral injection well. The pressure responses are compared between the corresponding reference vertical injection well configuration (blue curves) for the three reservoir permeability scenarios.

It is apparent from Figure 32 that the pressure response in horizontal well is equivalent to vertical wells at any location away from the injector. The deviation seen for the 8 mD reservoir permeability (left column of panels) between the horizontal and vertical wells during the last 2 injection periods can be attributed to boundary effects.

6.4 Summary

This report utilized a 2-D radial model to generate synthetic pressure responses following CO_2 injection into a depleted oil reservoir to evaluate the performance of horizontal well configurations for CCUS projects. The study successfully demonstrates the improved performance obtained by employing horizontal well configurations. A simple empirical correlation is obtained to estimate horizontal well performance in a given reservoir as a function of the half-length of the lateral. It is noteworthy that horizontal wells result in higher injectivities while resulting in equivalent pressure response to conventional vertical wells at any location away from the injector.

6.5 References

1. Ravi Ganesh P, Mishra S, Haagsma, Gupta N. Dynamic modeling to understand pressure response from oil production and CO_2 injection in a depleted pinnacle reef reservoir: Manual calibration using simplified resolution of reservoir heterogeneity. International Journal of Greenhouse Gas Control, Vol. 108, 2021.

2. Craft BC, Hawkins MF. Applied petroleum reservoir engineering. Englewood Cliffs, NJ: Prentice-Hall Inc. 1959.

3. Palacio JC, Blasingame TA. Decline curve analysis using type curves – Analysis of gas well production data. SPE paper 25909 presented at the 1993 Joint Rocky Mountain Regional/ Low Permeability Reservoirs Symposium, Denver, Colorado, USA. 26-28 April, 1993.

Chapter 7.

Field Data for Imaging Pressure and Temperature in the Subsurface using Tomographic Methods

Sanjay Mawalkar, Ashwin Pasumarti, Rick Pardini and Srikanta Mishra

Chapter 7. Table of Contents

Chapter 7. List of Tables

Chapter 7. List of Figures

7.1 Introduction

Geologic CO_2 sequestration and CO_2 enhanced oil recovery (EOR) have received significant attention from the scientific community as a response to climate change from greenhouse gases. Safe and efficient management of a CO_2 injection site requires spatio-temporal tracking of the CO_2 plume in the reservoir during geologic sequestration. To that end, as part of this research project, the Texas A&M team is developing robust modeling and monitoring technologies for imaging and visualization of the CO_2 plume using routine pressure/temperature measurements. Their approach is based on streamline-based technology which has proven to be effective and efficient for reconciling geologic models to various types of reservoir dynamic response. Specifically, they first extend the streamline-based data integration approach to incorporate distributed temperature sensor (DTS) data using the concept of thermal tracer travel time. Then, a hierarchical workflow composed of evolutionary and streamline methods is employed to jointly history match the DTS and pressure data. Finally, CO_2 saturation and streamline maps are used to visualize the CO_2 plume movement during the sequestration process.

This chapter describes the data that was assembled by Battelle to support the above-mentioned objectives. Various types of data that are of interest are listed below:

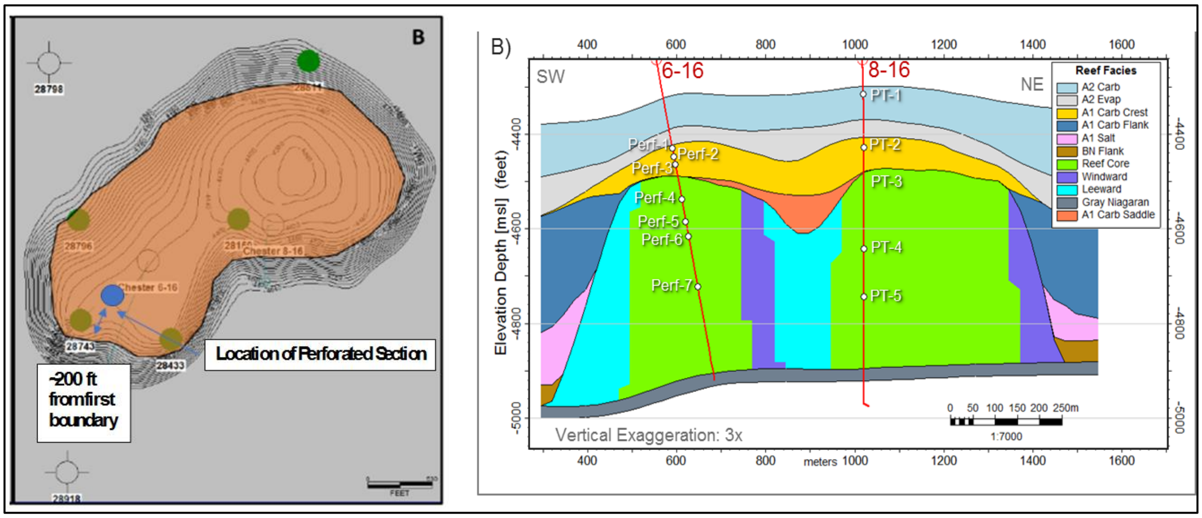
- Information about the geologic setting
- Oil production and CO2 injection history
- Bottomhole pressure and temperature data
- Distributed temperature sensing (DTS) data
- Numerical model developed by Battelle to integrate geologic, production and injection data
- Additional data types used for model validation (blind testing)

7.2 Project Overview and Data Sources

The Midwest Regional Carbon Sequestration Partnership (MRCSP) was established in 2003 to assess the technical potential, economic viability and public acceptability of carbon capture, utilization and storage [1]. This CO_2 injection project is part of the MRCSP and more than 1 million metric tons of CO_2 has been injected into Niagaran pinnacle reefs. The data of interest for this study are taken from the Chester 16 reef which is part of the Northern Niagaran Pinnacle Reef Trend. Chester 16 field was drilled and completed in the early 1970s and produced through the 1980s and 1990s. The reef has undergone primary recovery and some waterflooding. Currently, one injection well is used for reservoir fill-up with CO_2 prior to EOR operations, with one monitoring well which will be converted to a production well during the oil recovery period.

Figure 7-33 shows a map-view and three-dimensional view of the Chester 16 reef with one CO_2 injection well (Chester #6-16) and one monitoring well (Chester#8-

16). The Chester 16 reef includes two distinct reef cores close to one another. The reservoir in this reef is composed of two formations: A1 Carbonate and the Brown Niagaran formations. The CO_2 injection well #6-16 penetrates the reef complex at a high flank position in the southern reef core area and the monitoring well #8-16 penetrates the reef complex at a crestal position in the northern reef core area. The primary reservoir is the overlying A1 Carbonate (highly dolomitized high porosity zone along the crest of the reef). The A1 Carbonate is tight along the flanks of the reef, as is often the case and bounds the reservoir on all sides. The Brown Niagaran is a lower porosity reservoir with occasional fractures and/or dolomitic zones. A number of facies have been identified for the reservoir zones, as shown in the figure below. Also shown therein are the locations of the perforations and the multi-zone pressure and temperature sensors in the monitoring well



The left panel shows a contour map depth-surface of the Chester 16. The right panel shows a cross-section with reservoir facies, perforations and monitoring sensor locations in the two wells.

Figure 7-33. Map of the Chester 16 reef field showing well locations and structure of the Brown Niagaran.

The reservoir in Chester 16 reef was discovered and put into primary depletion in the 1960s for twenty years, after which waterflood was used for secondary production for another ten years and then the field was abandoned till 2017. CO_2 EOR was adopted in 2017 in this field for tertiary production. The field development history is summarized in Figure 7-34.

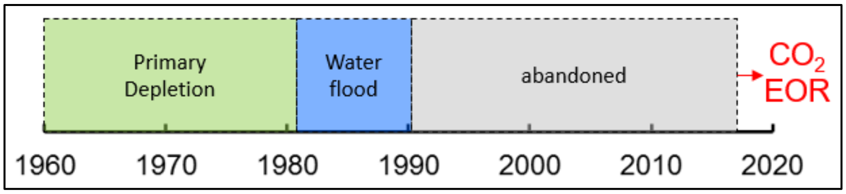
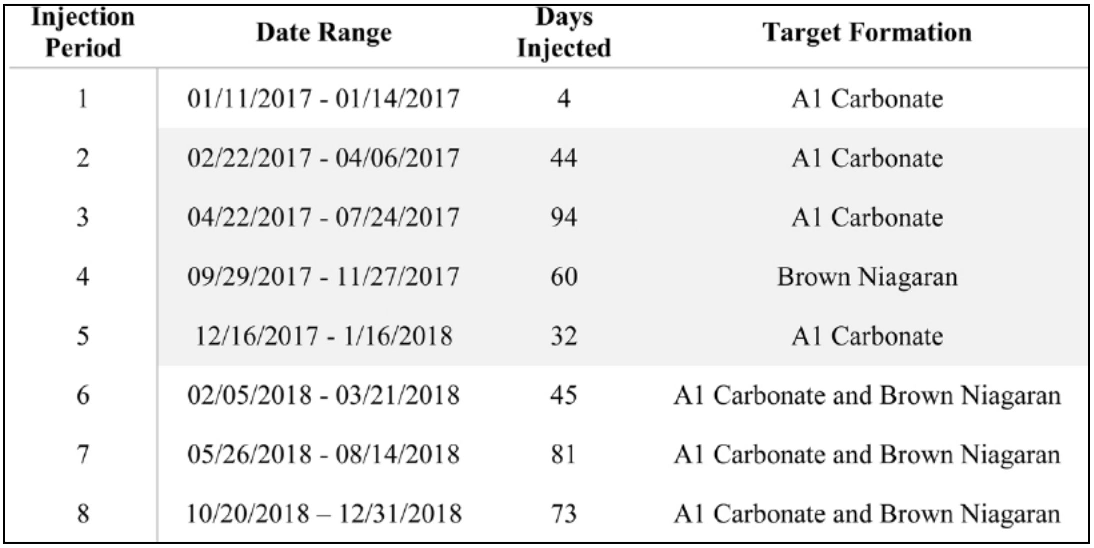


Figure 7-34. Chester 16 field development history.

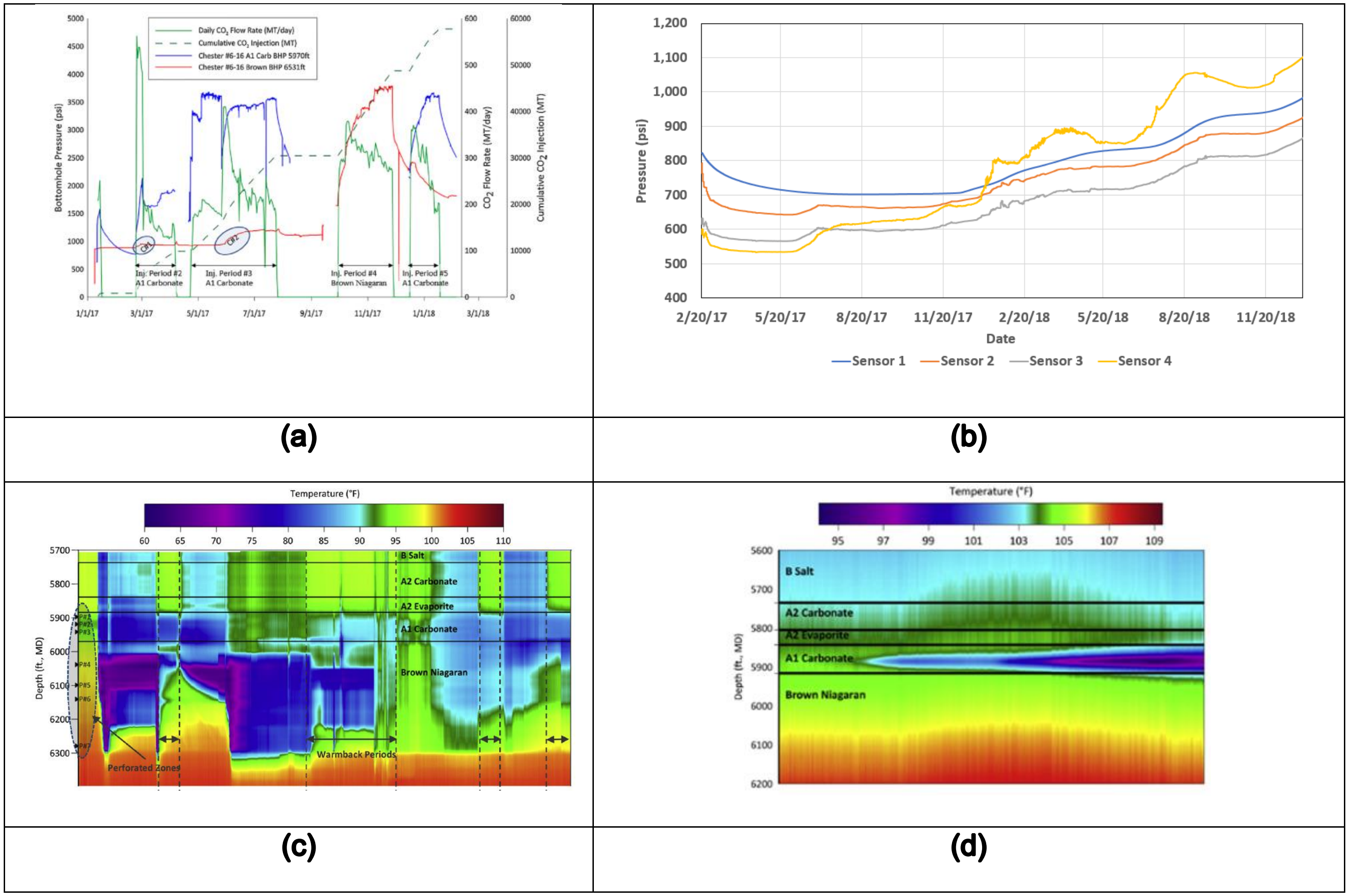
As mentioned above, the CO_2 injection started in 2017 and as of the end of 2018, ~101,000 MT of CO_2 has been injected in the Chester 16 reef. Table 7-14 shows the CO_2 injection periods and the intended injection formation(s).

Table 7-14. CO_2 injection history of Chester 16 reef [2]



Injection Period	Date Range	Days Injected	Target Formation
1	01/11/2017 - 01/14/2017	4	A1 Carbonate
2	02/22/2017 - 04/06/2017	44	A1 Carbonate
3	04/22/2017 - 07/24/2017	94	A1 Carbonate
4	09/29/2017 - 11/27/2017	60	Brown Niagaran
5	12/16/2017 - 1/16/2018	32	A1 Carbonate
6	02/05/2018 - 03/21/2018	45	A1 Carbonate and Brown Niagaran
7	05/26/2018 - 08/14/2018	81	A1 Carbonate and Brown Niagaran
8	10/20/2018 – 12/31/2018	73	A1 Carbonate and Brown Niagaran

In terms of observed data, this study utilizes a combination of pressure and temperature measurements, including bottom-hole pressure of injection well, distributed pressure measurements from four behind-casing sensors in monitoring well and DTS data of both injection well and monitoring well (Figure 7-35).



(a) **(b)**

(c) **(d)**

(a) Bottom-hole pressure of injection well; (b) Distributed pressure measurements from four behind-casing sensors in monitoring well; (c) DTS data at injection well; (d) DTS data at monitoring well

Figure 7-35. Observed dataset [2]

Both the wells in Chester-16 reef are also instrumented with a fibre-optic Distributed Temperature Sensing (DTS) system. The DTS data provides a rich source of time varying temperature data before, during and after CO2 injection. The data are collected vertically at 1-meter intervals, and temporally every hour. A waterfall plot of the temperature response at the injection well is shown below in Figure 7-36. Also shown is the bottom-hole pressure response at the injection well, and the injection rate history.

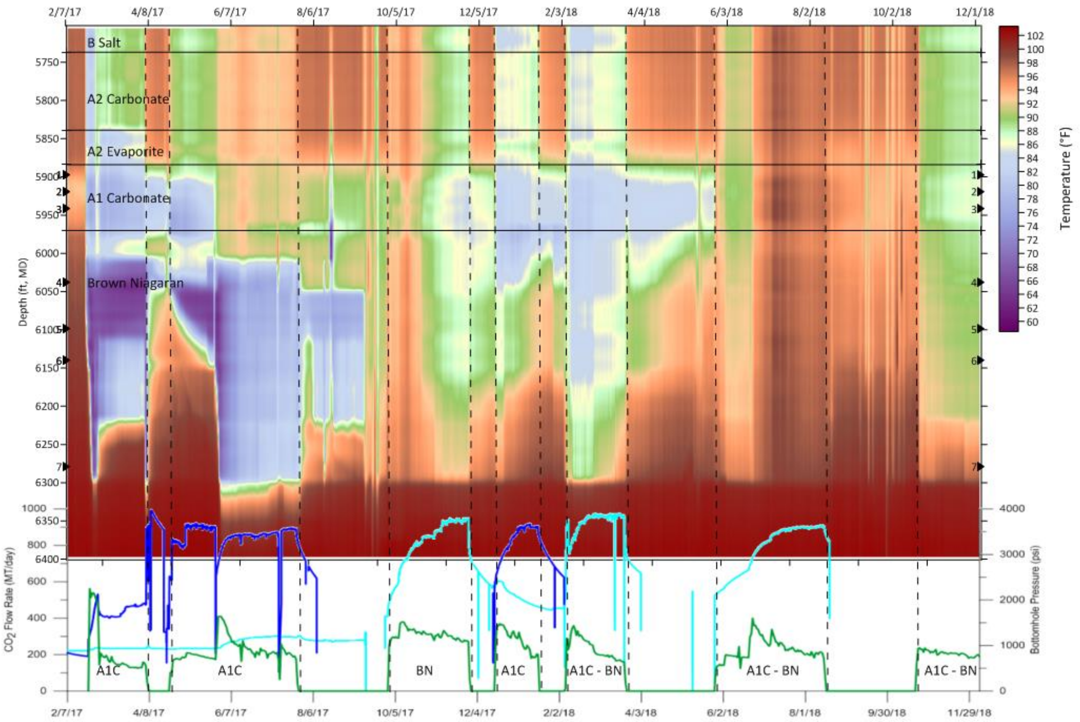


Figure 7-36. Master waterfall plot showing time-varying DTS response, along with bottomhole pressure data and CO_2 injection rate history at the injection well (Chester 6-16)..

For this project, Core Energy provided Battelle with operational data from the Chester 16 reef, comprising primarily of the Chester 6-16 injection well and the Chester 8-16 monitoring well. Both these wells are instrumented with a fiber-optic cable behind the production casing. This allows measurements of DTS temperature at every 1-m interval. The Chester 6-16 well is instrumented with memory gauges which can monitor bottomhole pressure and temperature in the A1 Carbonate or in the Brown Niagaran formation depending on the formation that is targeted for CO2 injection. Chester 8-16 well has behind-casing gauges installed at five depths which provide real time pressure and temperature data in the A2 Carbonate, A1 Carbonate and Brown Niagaran formations.

Mass flow rate measurements are available with a Coriolis flow meter attached to the Chester 6-16 injection well. Meanwhile, Chester 8-16 well underwent multiple transformations, starting out as a monitoring well until September 2019 when the well was perforated for CO2 injection in the A1 Carbonate formation. Later in May 2020, the Chester 8-16 well was perforated near the top of the Brown Niagaran formation for production purposes. As such, depending on the phase in which Chester 8-16 well operated, the data available includes bottomhole pressures and temperatures, flow rate of injected CO2, or the flow rate of produced gas, oil, and brine mixture during production phases. In November 2020, a new production well Chester 9-16HD1 was drilled to completion for producing oil from top of the Brown Niagaran formation. Table 7-15 below summarizes various data available from the Chester 16 reef.

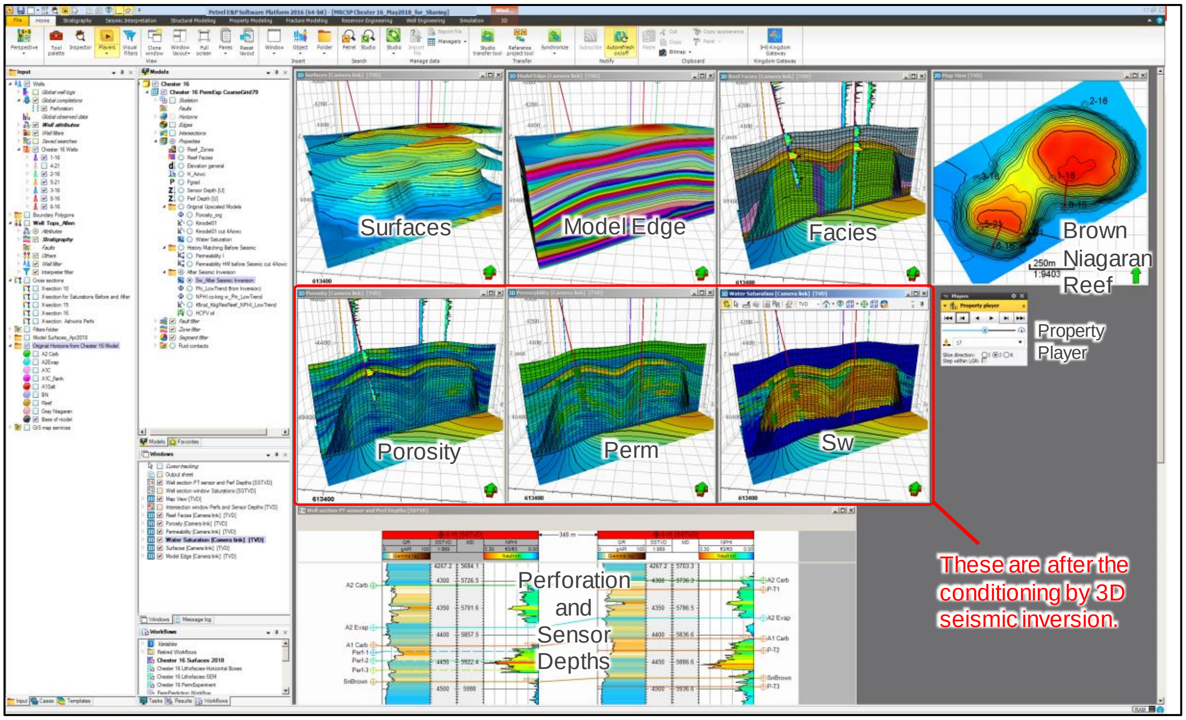
Table 7-15. Data Available from Chester 16 Reef

Well	Data Type	From Date	To Date
6-16 Injection Well	Bottomhole Pressures (BHP) and Temperatures (BHT)	Jan 2017	Jan 2020
6-16 Injection Well	DTS data to 6400’ MD	Feb 2017	Nov 2020
6-16 Injection Well	Injection Flow Rates (daily)	Jan 2017	Oct 2021
8-16 Well	Behind-Casing Sensors Pressures and Temperatures at 5 depths (includes monitoring, injection, and production phases of the well)	Feb 2017	Oct 2021
8-16 Well	DTS data (includes injection and monitoring phases to 6400’ MD, and production phase 6150’ MD)	Feb 2017	Jul 2020
8-16 Injection Phase	Injection Flow Rates (daily)	Sep 2019	Oct 2021
8-16 Production Phase	Produced Gas Flow Rates (daily)	Apr 2020	Nov 2020
9-16 HD1 Production	Produced Gas Flow Rates (daily)	Nov 2020	Oct 2021

7.3 Model Description

In this section, we will demonstrate the base case simulation model, including the model description, initialization, and simulation of production phases before the CO_2 EOR process. Additional details for the model can be found in Mishra [3].

A detailed geologic model was built for Chester-16 as part of the MRCSP modeling efforts. This static earth model (SEM) has been modified for use in this project. Figure 7-37 shows some of the model features as represented in the PETREL earth modeling software.



These are after the conditioning by 3D seismic inversion.

Figure 7-37. Chester-16 SEM features in the PETREL model.

Figure 7-38 (a)-(b) shows the permeability and porosity distribution of the base case model, with permeability ranging from 1e-10 md to 129 md and porosity ranging from 0 to 0.275. The 79-layer geologic model is discretized into 0.1 million grid cells, of which about 60,000 cells are active. Figure 7-38 (c) shows a side view of the model with two red arrows pointing to the two target formations of CO_2 injection: A1 Carbonate and Brown Niagaran, which are also the main formations of the reservoir.

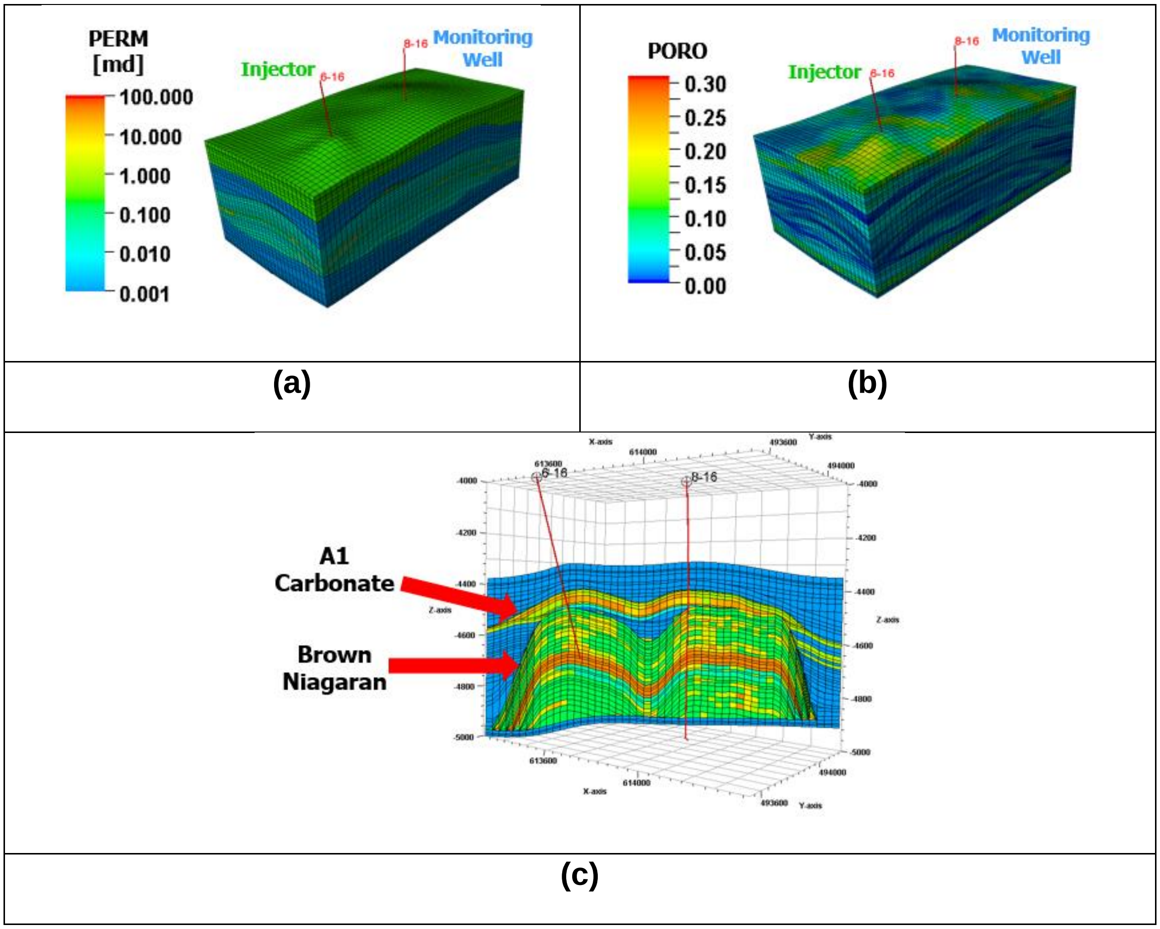
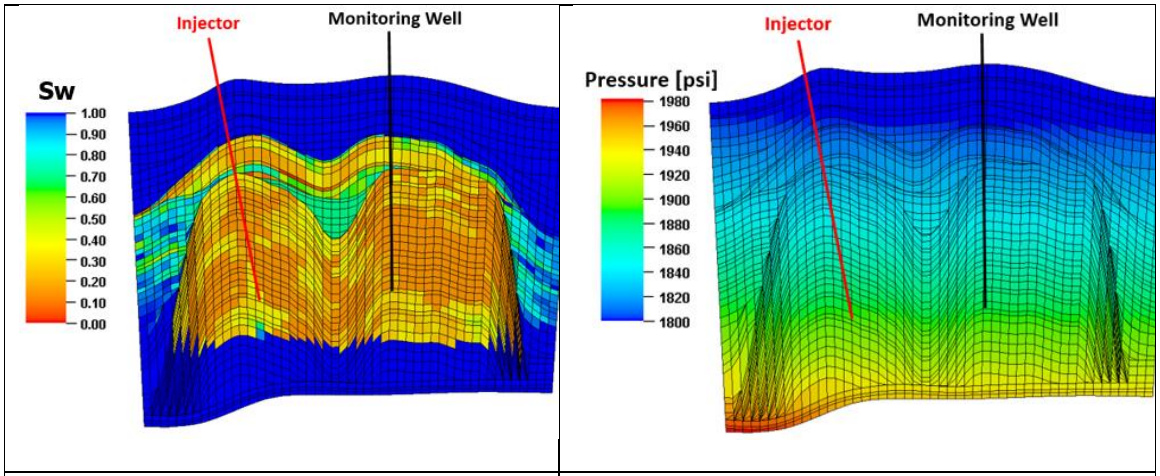


Figure 7-38. Base case model: (a) permeability distribution; (b) porosity distribution; (c) side-view denoting two target formation of CO_2 injection.

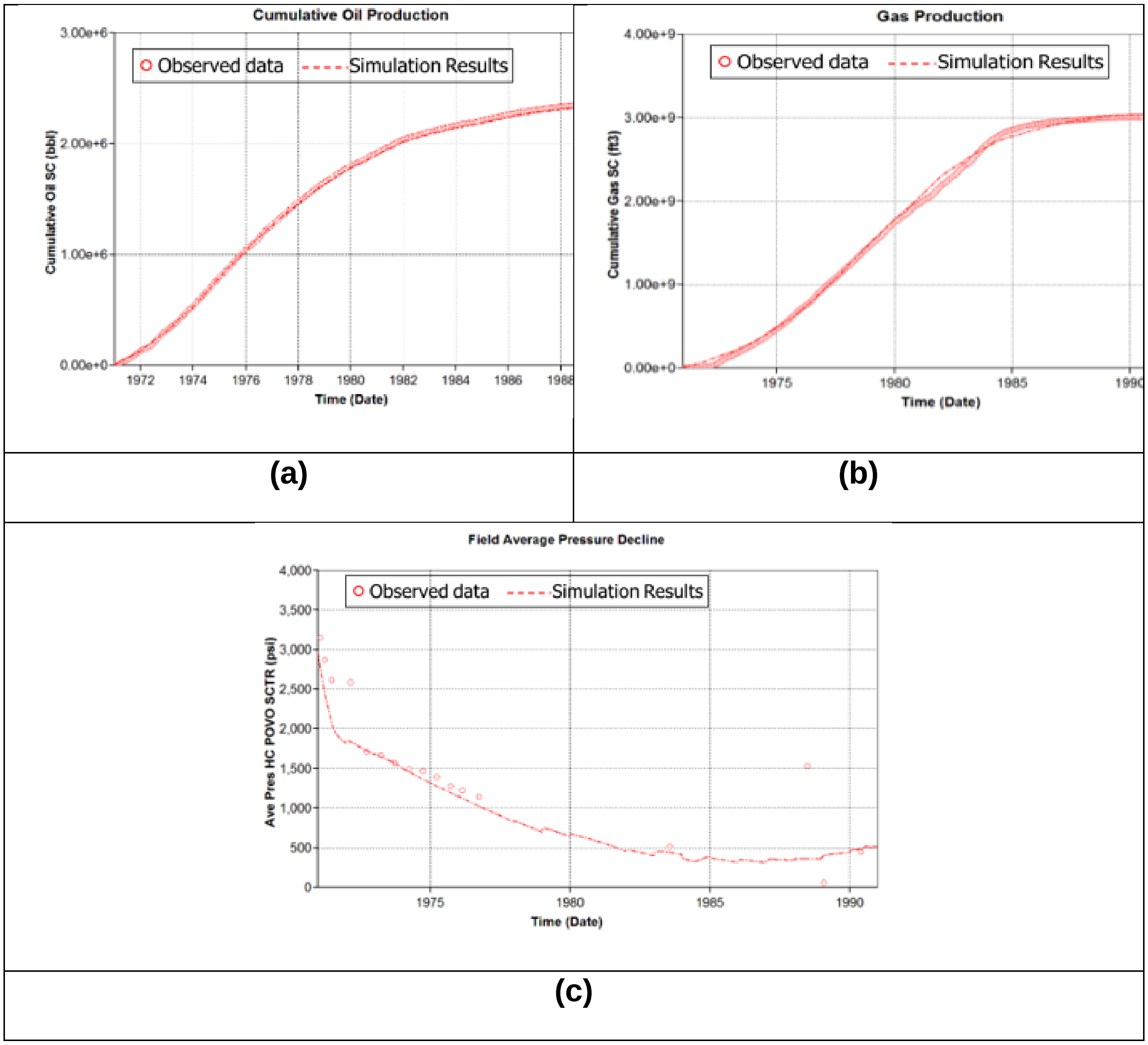
The fluid saturation of the base case model is initialized using seismic inversion results [3] and the pressure distribution follows the hydrostatic equilibrium. As shown in Figure 7-39, the average initial water saturation in the reservoir is about 0.2 and there is no gas cap in the reservoir.



(a) (b)

Figure 7-39. Base case model initialization: (a) Initial water saturation; (b) Initial pressure distribution.

In order to obtain the reservoir condition at the beginning of the CO_2 EOR process, simulation of previous production phases including primary depletion, water flooding and abandonment, was conducted. During this process, observed data such as cumulative oil production, cumulative gas production and average field pressure were matched, as shown in Figure 7-40.



(a)

(b)

(c)

Figure 7-40. History matching results for previous phases simulation: (a) Cumulative oil production; (b) Cumulative gas production; (c) Average field pressure

The simulation model verified by the observed data before the CO_2 EOR process is then used for the history matching of DTS and pressure data collected in the CO_2 EOR stage starting on January 1st 2017. The reservoir condition right before the CO_2 injection is shown in Figure 7-41. After 30 years of production and more than twenty years of abandonment, the average water saturation in the reservoir was reduced by around 0.05, and a large amount of gas exists at the top of A1 carbonate and Brown Niagaran formation. Since the reservoir is a closed system, the reservoir pressure was significantly depleted prior to the CO_2 injection.

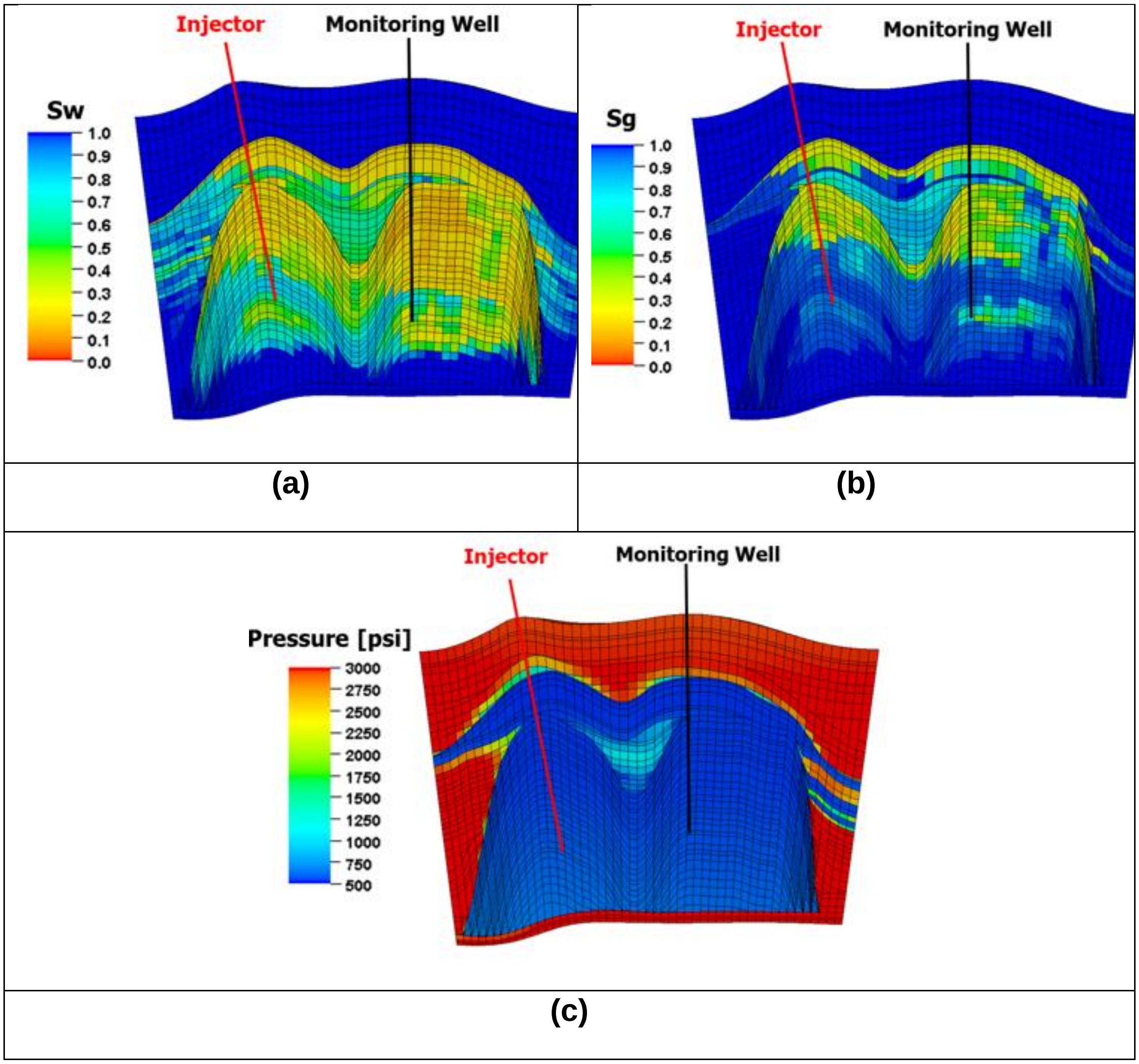


Figure 7-41. (a) Water saturation; (b) Gas saturation; (c) Pressure distribution on 1/1/2017.

7.4 References

1. Gupta N, Darrell P, Cumming L, *et al.* (2014). Testing for large-scale CO_2-enhanced oil recovery and geologic storage in the Midwestern USA, *Energy Procedia*, 63: 6393-6403.

2. Mawalkar S., Brock D., Burchwell A., Kelley M *et al.*, Where is that CO_2 flowing? Using Distributed Temperature Sensing (DTS) technology for monitoring injection of CO2 into a depleted oil reservoir, International Journal of Greenhouse Gas Control, Volume 85, 132-142, 2019

3. Mishra S., M. Kelley, A. Haagsma, P. Ravi Ganesh, J. Main, V. Smith, A. Pasumarti, S. Raziperchikolaee, A. Burchwell, (2020). *Numerical modeling of primary production and* CO_2 *injection performance in depleted oil fields in the Michigan Northern Pinnacle Reservoir Trend*, Final Topical Report, Midwest Regional Carbon Sequestration Partnership, Columbus, OH.

Chapter 8.

Machine Learning Based-Analyses for Time-Dependent Injection-Production Data

Po-Hsu (Allen) Chen, Brandon Hill and Srikanta Mishra

Chapter 8. Table of Contents

Chapter 8. List of Tables

Chapter 8. List of Figures

8.1 Introduction

The goal of this project is to characterize reservoir connectivity and optimize performance in injection-production systems using data-driven models. Several statistical, machine learning, and deep learning models were investigated for capturing the impact of time-dependent injection rates on the corresponding production rates. This report focuses on the comparison and summarization of the performance of those data-driven models. Two types of analyses were performed:

1. Time series forecasting – use of a model to predict future production rates based on previously observed production and injection rates,
2. Regression forecasting – use of a model to predict future production rates based on injection rates alone.

The key difference between the two analyses is whether the historical production rates are used or not as predictors. Section 8.2 shows the exploratory data analysis results, and Section 8.3 and Section 8.4 introduce data-driven models for time series forecasting and regression forecasting, respectively. The summary tables of results from different models can be found in Section 8.5, and Section 8.6 provides a discussion of the results.

8.2 Exploratory Data Analysis

8.2.1 Data Description

In this synthetic dataset, water was injected into 5 injection wells every other day, and the corresponding oil and water from 4 production wells were measured. For each production or injection well, there are 365 time steps (data points) after removing the first one, where all values are 0. Instead of modeling production oil and water rates, this report uses the *total production rate* (oil plus water) and the *water ratio rate* (water divided by total production), which provide the same information. **Figure 8-42** shows the distributions of injection, total production, and water ratio rates, where T1~T4 represent total production rates at production wells 1 to 4, WR1~WR4 are water ratio rates at production wells 1 to 4, and I1~I5 are injection rates at injection wells 1 to 5.

From **Figure 8-42**, we observe that daily injection rates seem to have a constant mean with noise. Furthermore, production wells seem to have two natural groupings: wells 1 and 4 are in one group, in which the total production rates are higher and fluctuating due to the changes of the injection water, while wells 2 and 3 are in another group, where the total production rates are much lower but steady. The steady total production rates from production wells 2 and 3 have raised some concerns, because production from these two wells seem to be totally independent of the volume of injected water. We also saw an increasing trend in production wells that may indicate there were some cumulative effects because the amount of the water injected overtime was relatively constant.

In this report, the first 275 data points were treated as the training data and the remaining data points were treated as testing data. **Figure 8-43** shows the boxplots for the rates of training and testing sets on each well (after converting oil and water rates to total production rate and water ratio rate). There are some potential outliers in the training data, all of which came from the first couple of dates. The analyses in the remaining sections did not remove those potential outliers, because more information is needed to determine whether they are true outliers. Another potential issue is that at most wells the testing data have higher rates than training data. Several data-driven models require scaling data to a 0-1 range because this helps in numerical calculation. But when there is a significant difference between training and testing data, these data will not be scaled separately.

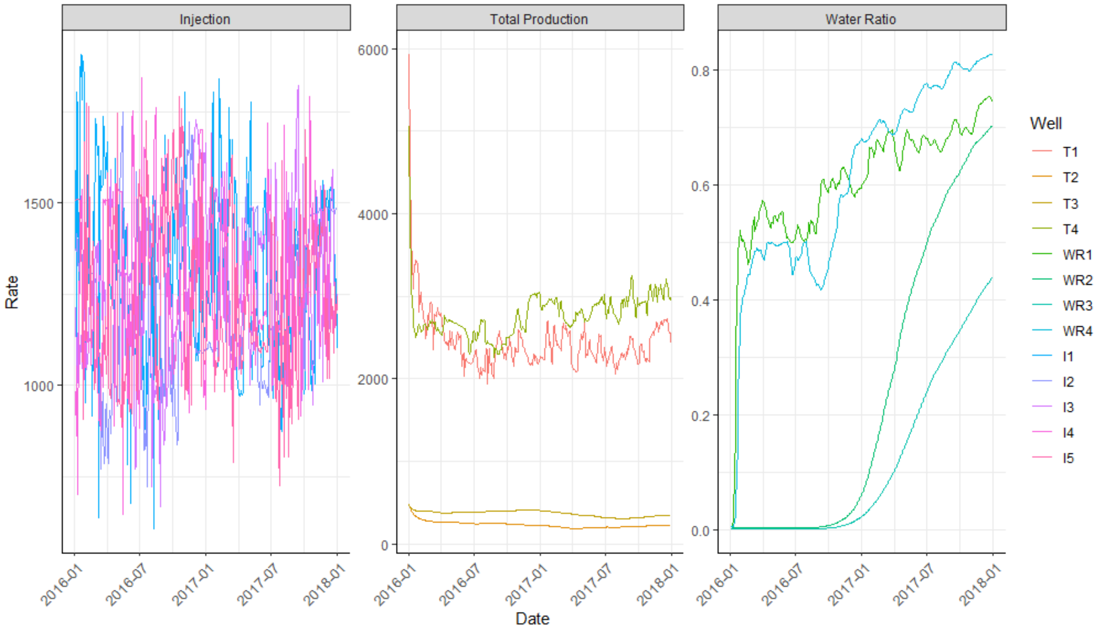


Figure 8-42. Comparison of injection, total production, and water ratio rates.

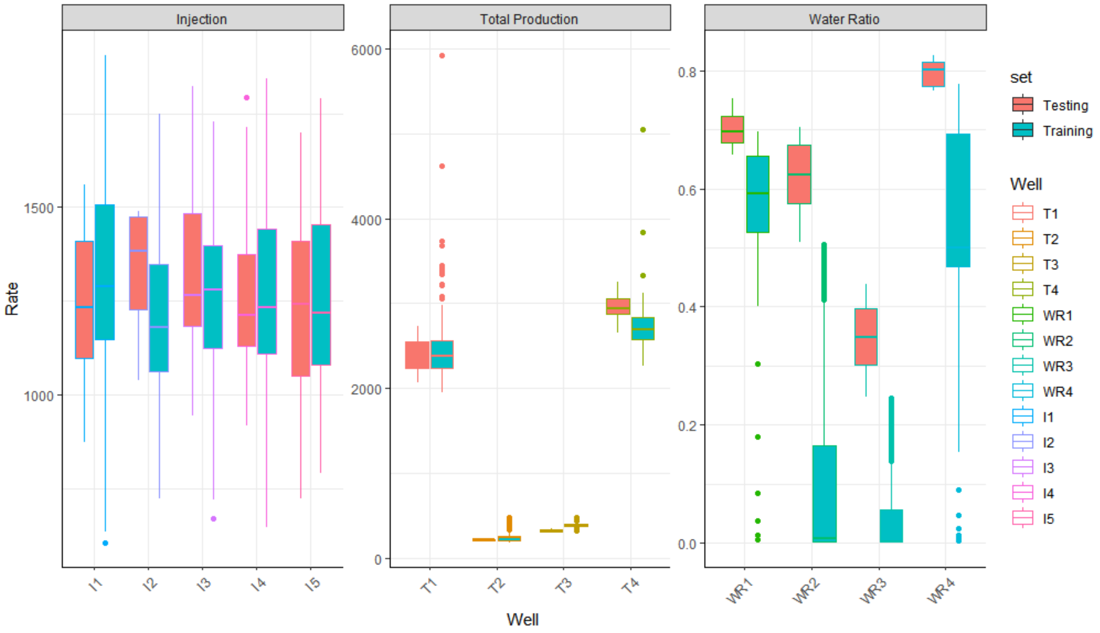


Figure 8-43. Boxplots for training and testing sets on each well

8.2.2 Cross-Correlation Analysis

Based on fluid-flow mechanics, there would be a time lag between injection and production as effects of the injection ripple through the reservoir to impact other wells. In order to visually evaluate the correlations between injection rate at each of the five injection wells and oil production at each of the four production wells, we created the cross-correlation plots shown in **Figure 8-44**. To produce these plots, we calculated the correlation between oil production on day *t* and injection rate on day (*t – Lag*), where *Lag* goes from 0 to 100. If there existed a strong correlation between injection and production rates, these plots would peak at/toward the beginning of the time series, and decay with increased lag time. Unfortunately, we see no such patterns in the cross-correlation plots, as most of the cross-correlations are relatively constant over lag time. In addition, we note that this relatively constant correlation is small, reaching a maximum of about 0.35 at best. We conclude that there is no significant correlation between injection rate and oil production at any time lag of any of the wells, and that most of the change in oil production over time cannot be attributed to daily changes in injection rate. There may still exist some relationship between oil production and injection rate, but the relation would be more complicated than a simple daily effect.

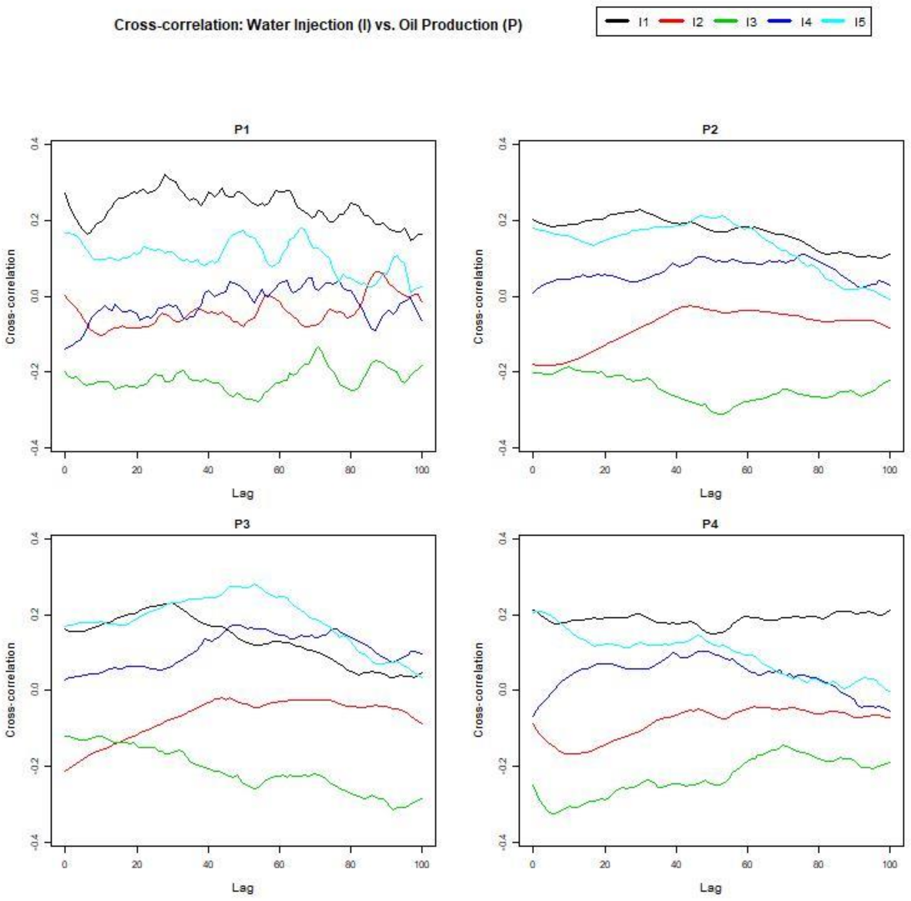


Figure 8-44. Cross-correlation between current oil production (one plot for each of the four wells) and water injection on previous days (injection wells are indicated by the color of the lines).

We see similar results in the analogous cross-correlation plots of injection rate at each of the five injection wells, and water ratio at each of the four production wells. Thus we also conclude that most of the change in water ratio over time cannot be attributed to daily changes in injection.

8.3 Time Series Forecasting

In this section we focus on time series forecasting (with Section 8.4 focusing on regression forecasting). We begin with brief introduction of models that were investigated. Next, we cover how to identify the optimal modeling approach in terms of prediction accuracy and model explainability, and then highlight potential issues of those models.

Models in this section used both production and injection rates at previous time points as predictors for future production rates. The window size in models determines how many production and injection rates of previous time points were used in prediction. For example, if the window size is 1, then today's injection rate and yesterday's production rate were used to predict today's production rate. Some models in this section may have slightly different numbers of training/testing data because of the different sizes of window applied (in which multiple recent time points were used to make a prediction for the next time point). This does not have any impact on the measurement of prediction accuracy, because root of mean square error (RMSE) (i.e., the average error across all points) is used in the evaluation of the performance across all models.

For time series forecasting, we investigated the following four models:

1) long short-term memory networks (LSTM) [1],
2) Gaussian process (GP) [2],
3) artificial neural networks (ANN), and
4) linear regression models.

We also compared fitting an independent model for each output or a model with multiple outputs and applying different window sizes. Note that since these are time series data, technically the future production rate should be predicted sequentially that only one time point will be predicted, and the predicted value will be used to as an input to predict the next time point. However, since it is challenging and time-consuming to add this feature to all the models that were compared, only the results using the observed value as predictor will be shown in this report.

8.3.1 Long Short-Term Memory Networks

LSTMs are a special kind of recurrent neural network (RNN) that work tremendously well on a large variety of problems and are now widely used. One popular application of LSTMs is in time series forecasting because its RNN structure with a short-term memory has the natural capability to make predictions based on historical data with lags of unknown duration between important events in a time series. In a LSTM network, three gates are present as shown in Figure 8-4 [3]:

1. Input gate – discover which value from input should be used to modify the memory.
2. Forget gate – discover what details to be discarded from the block.
3. Output gate – the input and the memory of the block is used to decide the output.

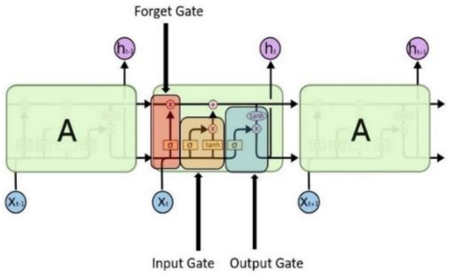


Figure 8-45. LSTM

Several LSTM models with different structures were investigated. For example, we started from the “independent models” for each single output (e.g., total production rate at well 1) at a time without considering other outputs’ impact (e.g., total production at well 2, 3, and 4). Then, we expanded the LSTM model to handle “multiple outputs”, i.e., all 4 total production rates were used to train a LSTM model. We also tested the impact of different window sizes.

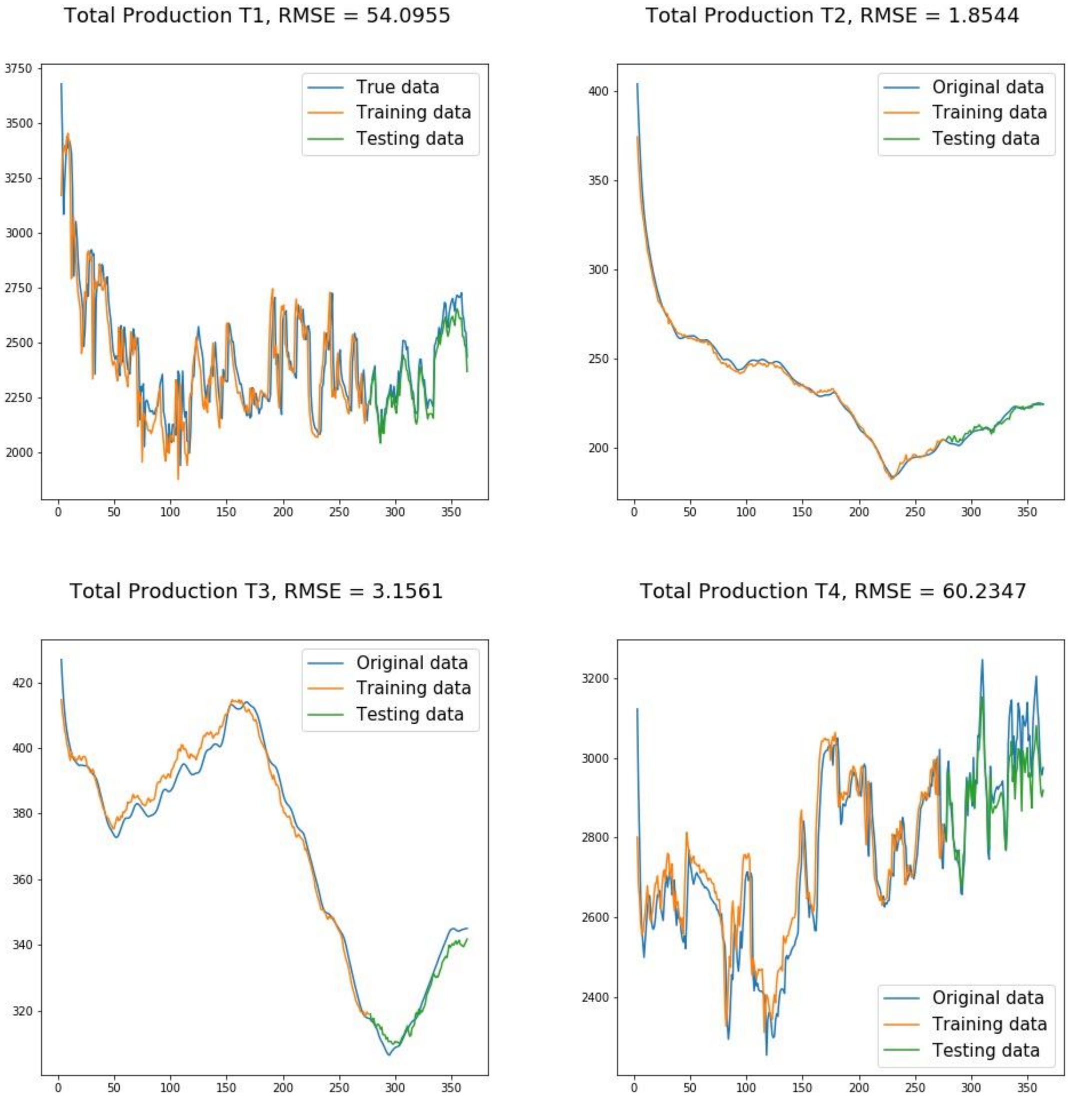


Figure 8-46. Predictions from the LSTM model with window size 3 and multiple outputs for total production.

The results from a LSTM model with multiple outputs and window size 3 are shown in Figure 8-46, because in general it produces the minimal RMSE scores across most of outputs. Results from other trained LSTM models can be found at Table **8-19** in Section 8.5. Generally, LSTM models can capture the fluctuating trend well if previous production rates were used in prediction, except for the first few time points in the training data (as mentioned in Section 8.2, those points would be the potential outliers) and the smoothness of the predicted values in time for wells 2 and 3. The large variation in predictions may imply that the LSTM models are too complex for this set of data. However, even though the prediction performance of LSTM models is reasonable, it is relatively difficult to characterize the connectivity between injection and production wells based on LSTM models because of their complicated structure.

8.3.2 Gaussian Process Models

GP models have a wide range of applications but have characteristics that make them especially useful for computer experiments. For example, GP models have the flexibility to mimic complex simulators because of their semi-parametric properties; that is, a GP model allows the user to specify the overall mean structure to fit the global trend, while at the same time, the model is capable of capturing local variation using a covariance (correlation) function. In addition, the empirical best linear unbiased predictors (BLUPs) of GP models interpolate the data at observed input settings, which is an important property for modeling deterministic computer simulators. GP models are also popular in time and spatial data analyses due to the fact that GP models have an assumption that two (geographically) close points will have similar output behavior. For more technical details of GP models, refer to [4].

Even though some approaches for training GP models with multiple outputs have been developed (see [5] as an example), most approaches are still immature, and no well-written package in software has been developed for the implementation of those models. Thus, only GP models with a single output were considered in the report, and the MATLAB *MPErK* [6] package was used to predict production rates and calculate sensitivity indices.

Since window size = 3 provided the optimal results for LSTMs as shown in Subsection 8.3.1, the GP models were also trained with window size = 3 to predict future production rates, facilitating easier comparison between the models . **Figure 8-47** shows that GP models also provide a decent result in terms of prediction accuracy for production wells 1 and 2. For production wells 2 and 3, GP models seem to have the capability to capture the overall smooth trend in the data but for production well 3 for a range of data the predictions from GP model diverged substantially from the observed values. For production well 4, the GP model captured the fluctuations in the response, but did not predict as well as the LSTM models.

Even though the prediction accuracy from GP models may not be as good as LSTM models, one advantage of GP models is the closed-form expression that allows the calculation of several diagnostic measures to help understand the quality of the model fit. For example, sensitivity analysis for GP model is an approach to determine how input variables affect the outputs. The most popular sensitivity analysis is based on an ANOVA-type decomposition [7]. Sensitivity indices are values between 0 and 1, in which an input variable with a large value indicates that it has a strong impact on the output. The differences between main and total effects in sensitivity indices can be interpreted as the sum of interaction effects for that input variable. Another well-known sensitivity analysis method is based on the partial derivatives of the outputs of the model with respect to each input, see [8] for example.

For demonstration purposes, consider the GP models with a window size = 1, which have fewer variables than the window size = 3 models described above. The sensitivity indices for these models are shown in **Table 8-16**.

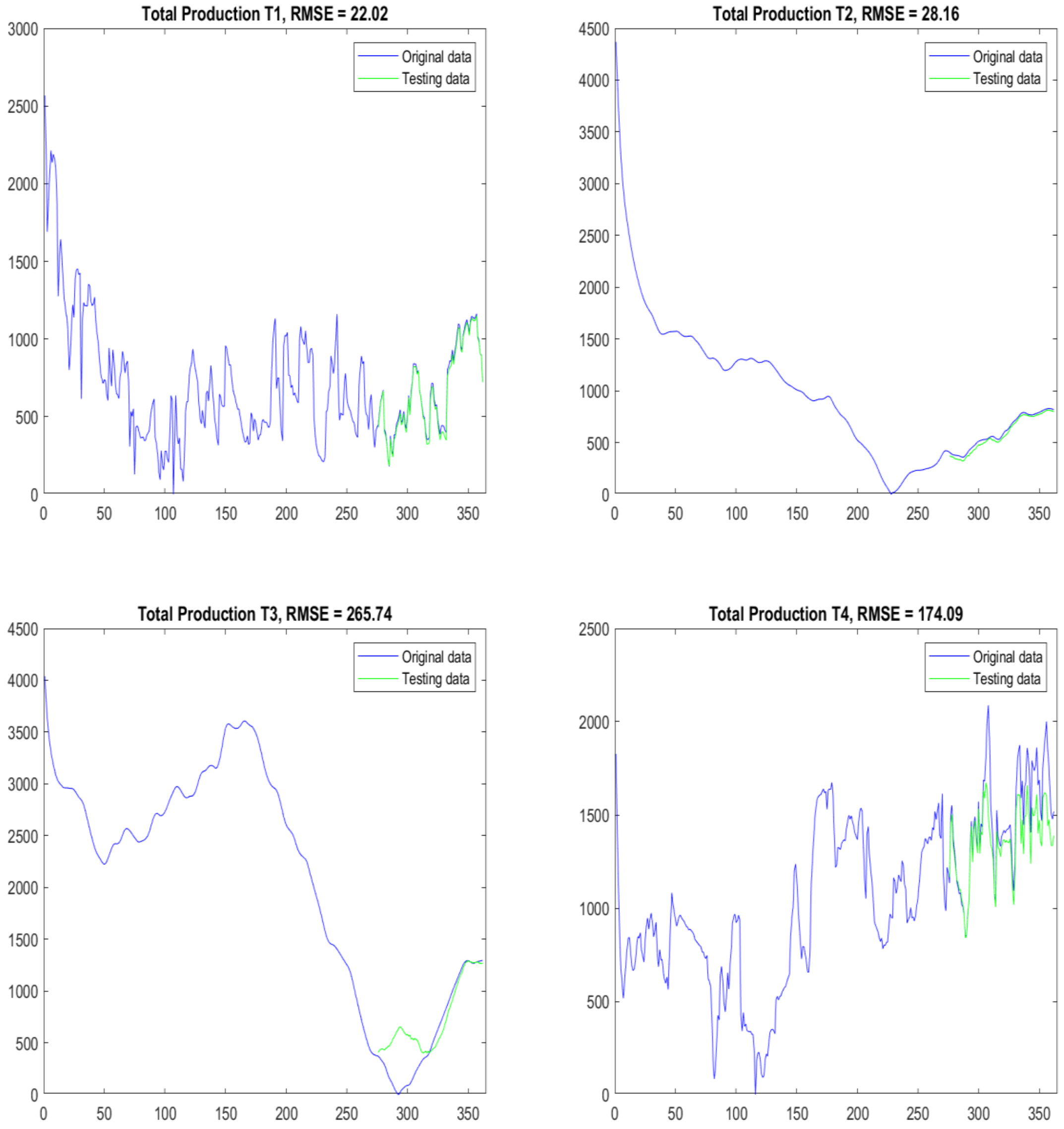


Figure 8-47. Predictions from the GP models with window size 3 for total production rates.

Sensitivity analyses indicates that the total production rate on all wells heavily depend on the previous total production rate, P(t-1). For T1 and T4, injection rates at wells 1 and 3 also play an important role, respectively. However, total production at wells 2 and 3 seems to only depend on previous total production rates but statistically unrelated to the injection rates, as observed in Section 8.2.

In addition, main effect plots can provide a general idea about how input variables impact the output when the effects from all other variables were integrated out. For example, **Figure 8-48** shows the main effect plot for the total production rate at well 1, in which variables x1~x6 are corresponding to P(t-1), I1, …, I6 in Table 8-16 and we can see the previous production and injection well 1 have a positively correlated linear impact on the total production at well 1. Other GP model results can be found in Section 8.5.

Table 8-16. Sensitivity indices for GP model with window size = 1. P(t-1) denotes the production rate at previous time point, I1~I5 represent injection wells 1 to 5, T1~T4 represent total production rate at wells 1 to 4.

Prod.	Effect	P(t-1)	I1	I2	I3	I4	I5
T1	Main	0.50	0.28	0.05	<0.01	<0.01	<0.01
	Total	0.64	0.35	0.21	<0.01	0.01	0.01
T2	Main	0.99	<0.01	<0.01	<0.01	<0.01	<0.01
	Total	0.99	<0.01	<0.01	<0.01	<0.01	<0.01
T3	Main	0.99	<0.01	<0.01	<0.01	<0.01	<0.01
	Total	0.99	<0.01	<0.01	<0.01	<0.01	<0.01
T4	Main	0.73	<0.01	<0.01	0.21	<0.01	<0.01
	Total	0.78	0.01	<0.01	0.26	0.01	<0.01

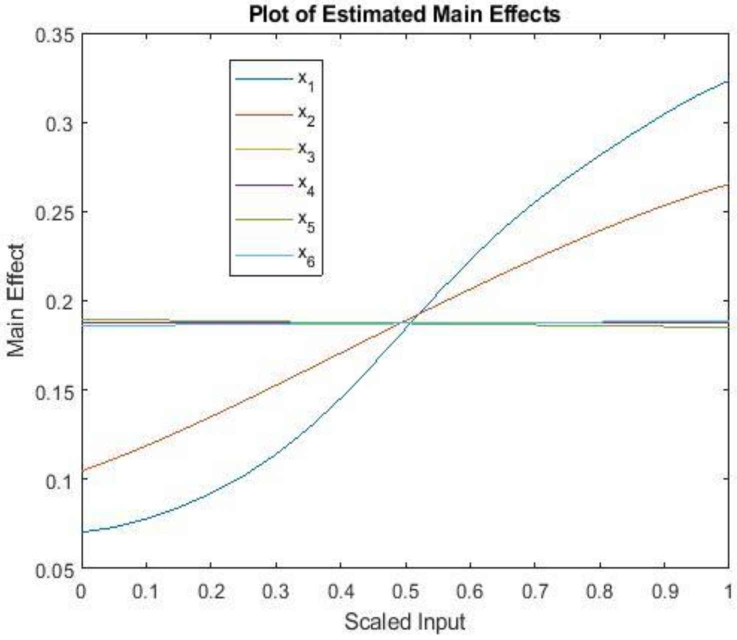


Figure 8-48. Main Effect Plot of a GP model.

8.3.2.1 Main Effect Plot of a GP model

Another model we investigated for this study is the artificial neural network (ANN), which is one of the main tools used in machine learning. As the "neural" part of their name suggests, they are brain-inspired systems which are intended to replicate the way that humans learn. Neural networks consist of input and output layers, as well as a hidden layer consisting of units that transform the input into something that the output layer can use. **Figure 8-49** provides an example of ANN with 5 inputs, 4 outputs, and one hidden layer with 15 neurons. ANNs have been successfully used for variety of tasks and fields including waterflooded reservoirs; see [9] for an example where the authors built an ANN model to estimate the interwell connectivity between injection and production wells, and [10] for a case study and [11] for analyzing cyber-physical petroleum systems.

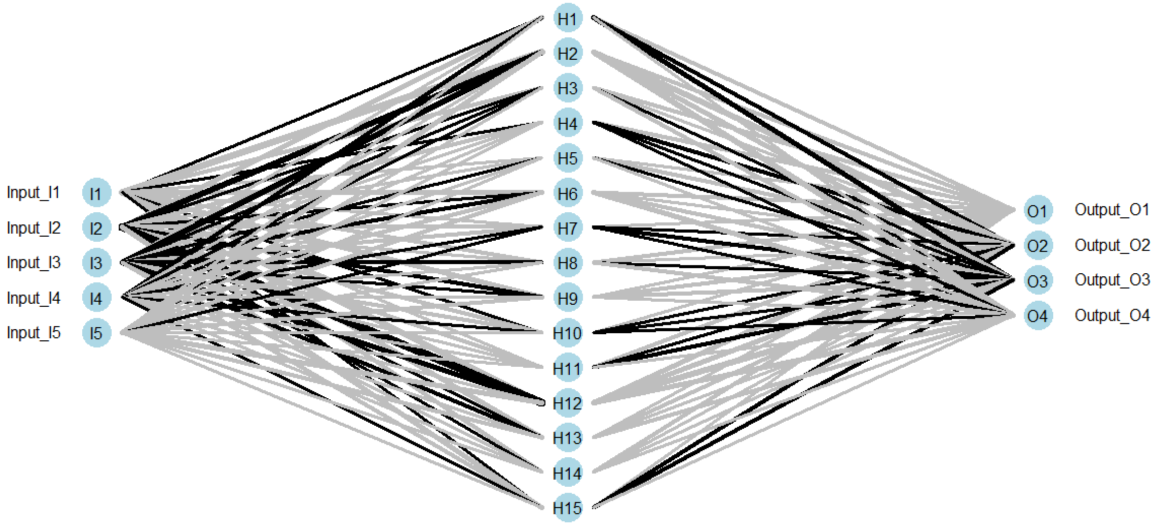
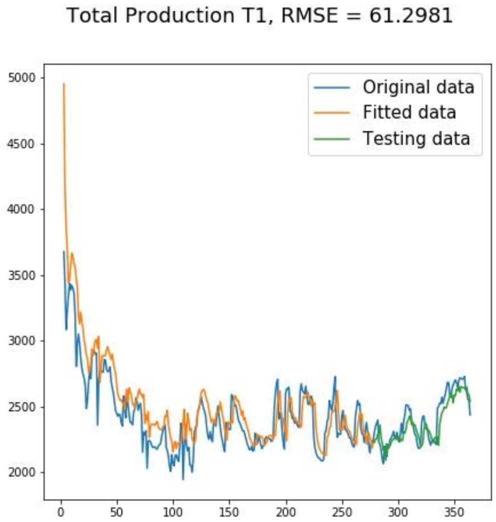
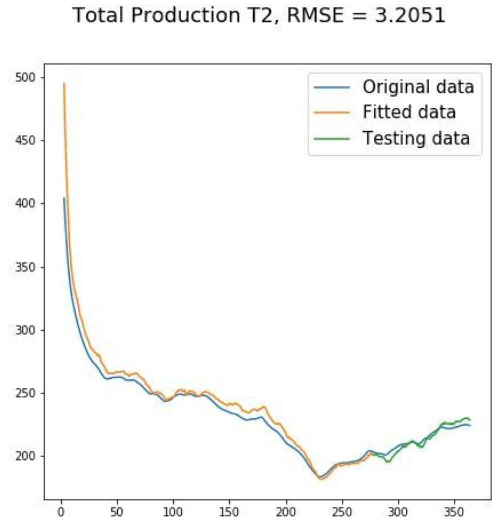


Figure 8-49. An example of an ANN structure with 5 inputs, 4 outputs, and a hidden layer with 15 neurons.

A multivariate outputs (4 outputs) ANN model with a hidden layer of 15 neurons was trained to predict production rates. As with the other approaches, the first 275 data points were used to train the model to predict the remaining data points. Moreover, a window size of 3 was used for consistency with LSTM and GP model results. Predictions from the ANN model are shown in **Figure 8-50**. Generally speaking, the performance of ANN and LSTM models are equally matched, with both of them providing a decent prediction result.

In addition to prediction accuracy, understanding the importance of variables from the trained models is one of the main goals in this project. To aid in the interpretation of neural networds, the R package *NeuralNetTools* [14] was used to visually interpret the trained ANN models. In Subsection 8.4.1 and Subsection 8.4.2, a visualization method and three variables importance algorithm are introduced for charactering reservoir connectivity. In Subsection 8.3.2 (sensitivity analysis based on GP models), a simplified model using only one previous time point production rate as a predictor was considered, following the same training process for the ANNs. Four independent ANN models were trained, all of which have 6 inputs (5 injection rates plus previous 1 time point production rate) and one hidden layer with 15 neurons and one output.


Total Production T1, RMSE = 61.2981


Total Production T2, RMSE = 3.2051

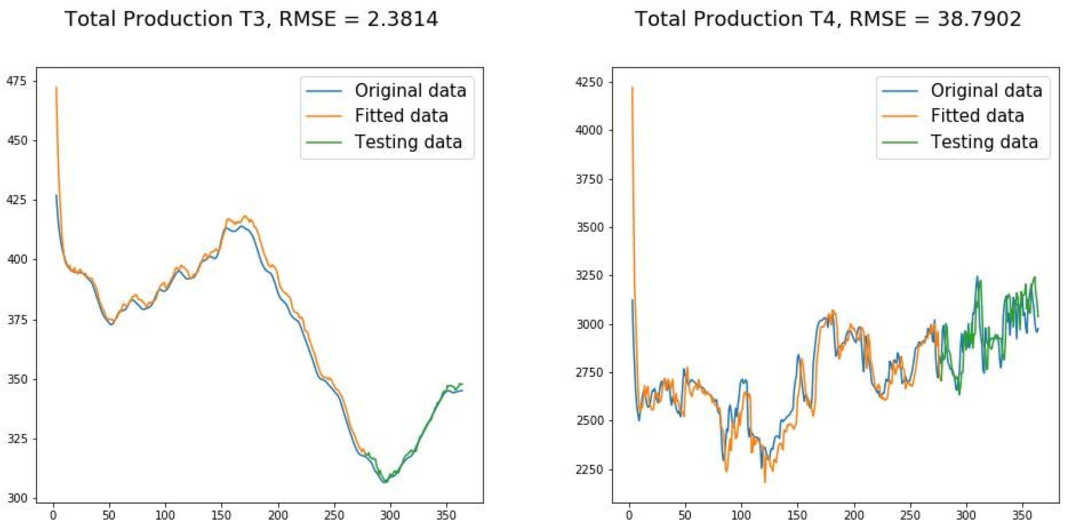


Figure 8-50. Predictions from the ANN model with window size 3 and multiple outputs for total production.

8.3.3 Neural Interpretation Diagram

A neural interpretation diagram (NID) is a modification of the standard conceptual illustration of the ANNs that changes the thickness and color of the weight connections based on magnitude and sign, respectively. Positive between layers are shown as black lines, while negative weights are shown as gray lines. Line thickness is proportional to the absolute magnitude of each weight. **Figure 8-51** shows the neural network architecture and the variation in connections between the layers for the total production at well 1, where the largest positive weight is the connection between injection well 1 and the hidden node 11. However, it is difficult to interpret given the amount of weighted connections in this network.

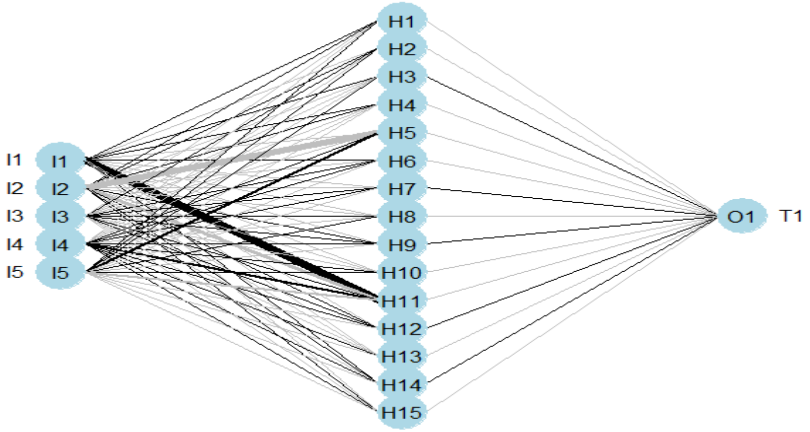


Figure 8-51. The NID plot for total production at well 1.

8.3.3.1 Variables Importance Analysis

As mentioned in Subsection 8.3.2.1, the connectivity between all input-output pairs of neural networks can be used to quantify the importance of variables. Three algorithms for evaluating variable importance are applied: Garson's algorithm for relative importance [14], Olden's connection weights algorithm [12], and Connectivity map approach [9].

8.3.3.1.1 Garson's Algorithm

Garson's algorithm calculates the summed products of all absolute weights specific to each input variable, and then those values are scaled relative to all other inputs. A value for each input variable indicates relative importance as the absolute magnitude from zero to one. However, this method is limited in that the direction of the response cannot be determined and only neural networks with one hidden layer and one output node can be evaluated. Thus, only single output ANN models were considered in this subsection for a fair comparison. Four independent ANN models were trained, all of which have 6 inputs (5 injection rates and total production rate at previous one time point) and one hidden layer with 15 neurons and one output. The relative importance of variables from the trained four independent ANN models using Garson's algorithm is shown in **Figure 8-52**.

8.3.3.1.2 Olden's Connection Weights Algorithm

The Olden's connection weights algorithm is a more flexible approach which calculates importance as the summed product of the raw input-hidden and hidden-output connection weights between each input and output node. Its advantage is the relative contributions of each connection weight are maintained in both magnitude and sign. There is a strong assumption behind this method – negative contributions are low connectivity in the reservoir, which may not be true for different applications. The importance of variables from the trained four independent ANN models using Olden's algorithm is shown in **Figure 8-53**. Note that the 4 faceted plots in **Figure 8-53** have different scales on the y-axis (some inputs will dominate the importance measure, making it difficult to read if the same scale across all four outputs was used).

8.3.3.1.3 Connectivity Map

The first step in the construction of a connectivity map is to calculate the contribution of each input using Olden's connection weights algorithm. Then, a connectivity map can be constructed by normalizing those contributions between 0 and 1 (0: lowest connectivity, 1: highest connectivity). The constructed connectivity map is shown in **Figure 8-54**. Note that both Olden's connectivity weights and connectivity map can be applied to ANN models with multiple outputs, while Garson's algorithm can be used in ANN models with single output. To do a fair comparison, four independent ANN models for each production well are trained for demonstration the results of different variables importance algorithms. Note that the connectivity map construction is using a function that we wrote in R based on the formula and the example in [9, 12]. The R package *NeuralNetTools* does not provide this function.

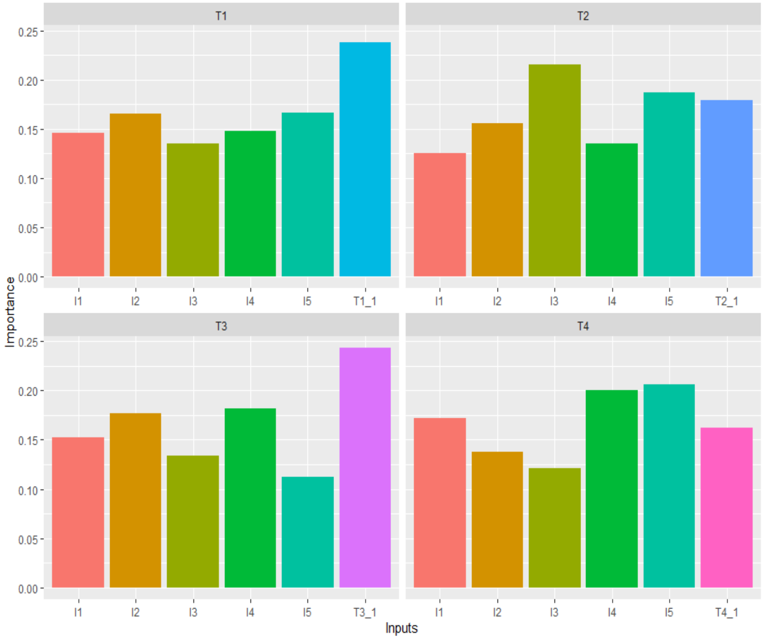


Figure 8-52. Variable importance of the 4 independent ANNs with previous production rate as a predictor using Garson's algorithm.

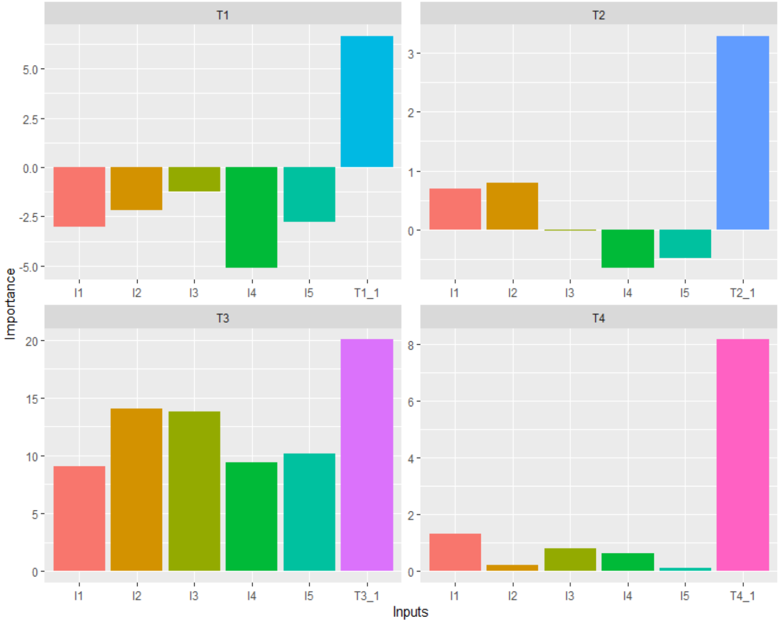


Figure 8-53. Variable importance of the 4 independent ANNs with previous production rate as a predictor using Olden's connection weights algorithm.

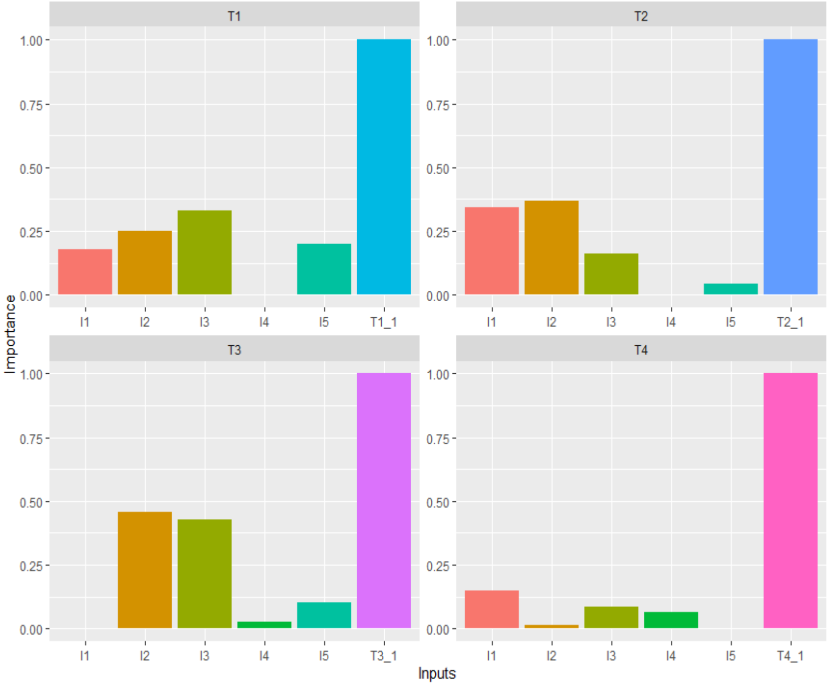


Figure 8-54. Variable importance of the 4 independent ANNs with previous production rate as a predictor using the connectivity map.

The relative importance of variables from the trained four independent ANN models using the three algorithms are shown in **Figure 8-52** through **Figure 8-54**, respectively, where T1_1,…T4_1 denote the previous total production rate for T1 to T4. All three algorithms show that the production rate from the previous time step played an important role, even though the calculation from Garson's algorithm did not represent the dominance of previous production rates as other algorithms or the results from sensitivity analysis of GP models. The negative contribution from Olden's connectivity weights algorithm (**Figure 8-53**) does not intuitively fit the physics-based insights for reservoir connectivity. Thus, we recommended to use the connectivity map to measure the contribution of each input variable (injection rate) for charactering reservoir connectivity as suggested in [9, 12].

Though the variable importance analyses (**Figure 8-52** through **Figure 8-54**) provide insights about the waterflooding dynamics in the reservoir and help understand the overall reservoir connectivity. However, it would be still a challenging task to interpret ANN models due to the fact that 1.) the most contributed input from ANN models is not necessary the inputs (injection rates) that we can adjust and optimize the production rates because they may not have a linear relationship, and 2.) the main contribution is the previous production rate, which does not provide much help in characterizing reservoir connectivity for understanding how injection rates affect the production rate in order to optimize the reservoir system. Using previous production rates as predictors improves the prediction accuracy, but it does not provide much insight into how injection and production wells are connected.

8.3.4 Linear Regression Models

In this subsection, a simple model, linear regression, is investigated for time series forecasting. The focus is on regression models with window size = 1 (only the production rate from the previous time step is treated as a predictor) because we do not expect that regression models can have better prediction accuracy than LSTMs or ANNs. Instead, the focus is to see whether linear regression models could provide more insight into reservoir connectivity without impacting prediction accuracy. Four independent linear regression models were fit for the total production rates at the four wells, all of which have 6 inputs (5 injection rates plus previous 1 time point production rate). **Figure 8-55** shows the prediction results for the four production wells and the corresponding RMSEs, see **Table 8-19** in Section 8.5 for more results of regression models.

The prediction accuracy from the above regression models are surprisingly almost as good as ANN and LSTM models. For example, for production well 1, the RMSE of the regression model is 37, while the RMSE of the ANN model with window size = 1 is 33, and for production well 4, the RMSEs of the regression model and ANN models are 88 and 67, respectively. In fact, for production wells 2 and 3, regression models outperform ANN and LSTM models, but this effect would be explained as occurring because the injection rates on the 5 injection wells almost have no impact on production rates.

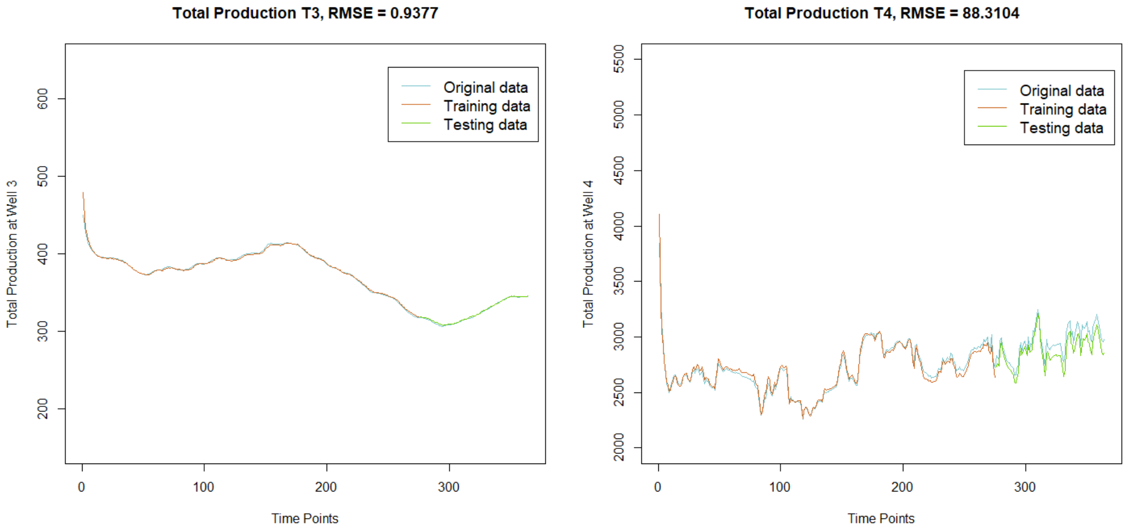


Figure 8-55. Predictions from the regression models with window size = 1 for total production rates.

P-values of coefficients of the regression models can be used to determine if the coefficient is significantly different from 0. However, p-values cannot be used to measure which variable is the most important one among those significant variables. Instead, Pearson's correlation coefficient is used, which is a measure of linear association between the two variables. In simple linear regression models, the square of Pearson's correlation, usually denoted as $R^{\wedge 2}$, is the proportion of the total variable of output that can be explained by the input variable. **Table 8-17** shows the square of Pearson correlation coefficients from simple linear regression models. The result in **Table 8-17** is consistent as the sensitivity analysis result based on GP models (see **Table 8-16**), both of which show that previous total production rate, P(t-1), is the most important variable for all production wells and injection rates at wells 1 and 3 are important for production wells 1 and 4, respectively.

Table 8-17. Square of Pearson correlation coefficients from simple linear regression models.

Production	P(t-1)	I1	I2	I3	I4	I5
Well 1	0.87	0.52	0.01	0.03	0.05	0.01
Well 2	0.99	0.06	<0.01	0.06	<0.01	0.03
Well 3	0.99	0.03	0.03	<0.01	<0.01	0.07
Well 4	0.87	<0.01	<0.01	0.50	0.01	<0.01

In this section, four different predictive models were evaluated in terms of prediction accuracy. All models provide decent prediction results when previous production rates were used as predictors. Unfortunately, one of the main goals in characterizing reservoir connectivity is to understand how injection rates affect the production rate in order to optimize the reservoir system. Even though using previous production rates as predictors improves the prediction accuracy, it does not help much in learning how injection and production wells are connected.

8.4 Regression Forecasting

Previous results showed that production rates from earlier time points were useful for predicting production in upcoming time points but including them in the model makes it difficult to understand the dynamics between injection and production wells. This section considers models that do not use past production data; even though it is expected that prediction accuracy will suffer, the hope is that simpler models may allow for better interpretation of those well dynamics.

8.4.1 Artificial Neural Networks

The previous section shows that ANNs seem to be a good choice for modeling reservoir connectivity, which is also consistent with the suggestion provided in [9 through 11, 13]. The trained ANN in this case has 5 inputs (5 injection rates), 4 outputs (4 production rates), and with one hidden layer with 15 neurons. The fitted results shown in **Figure 8-56** are reasonable for production wells 1 and 4 but quite poor for production wells 2 and 3.

As stated in Subsection 8.3.2.1 understanding the importance of variables from the trained models is one of the main goals in this project. Garson's algorithm, Olden’s connection weights algorithm, and connectivity maps (see Subsection 8.3.3.1) for the introduction of variables importance analysis) were used to analyze the importance of variables in the trained ANN models.

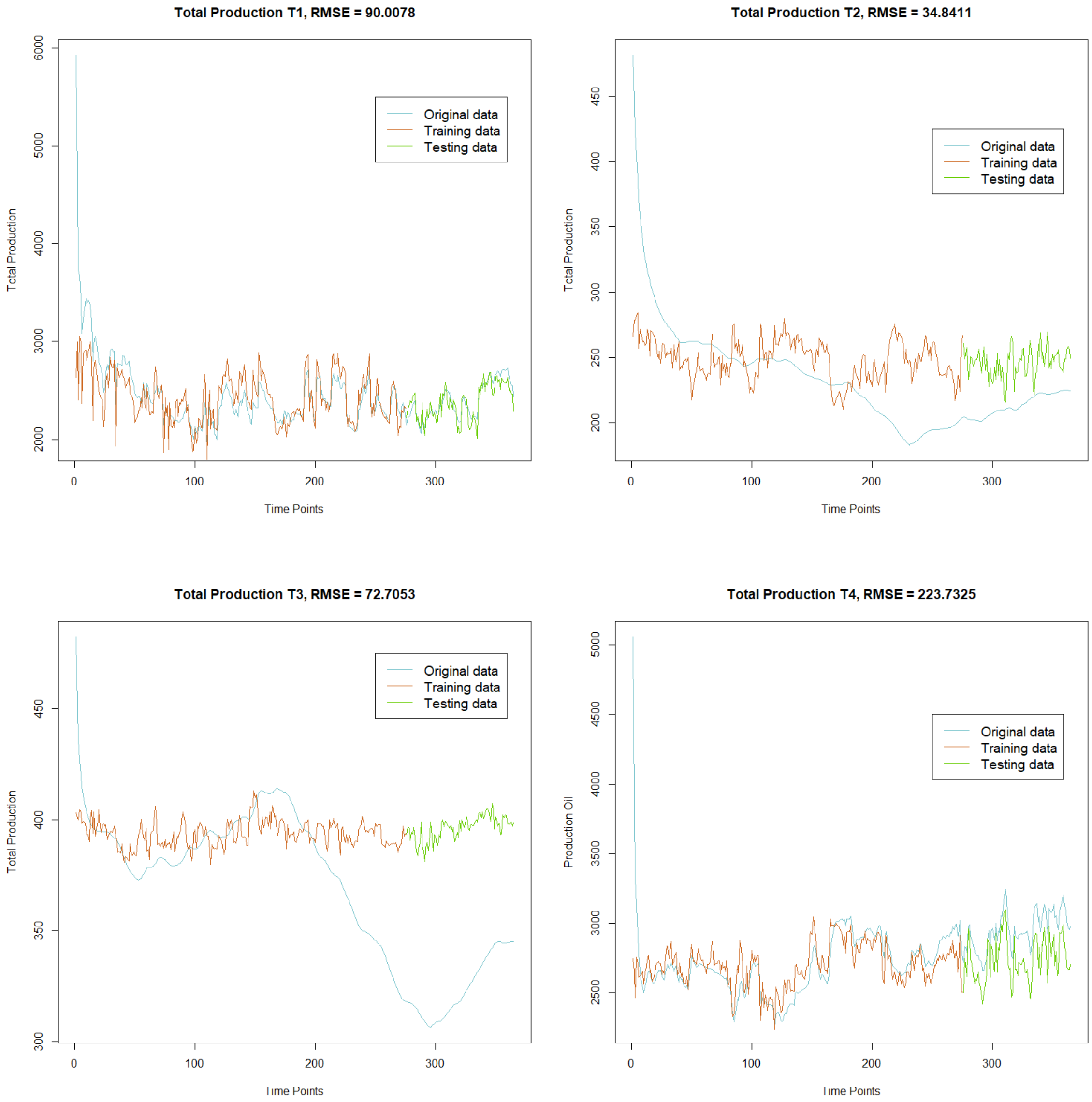


Figure 8-56. Predictions from an ANN model without using previous production rates as predictors.

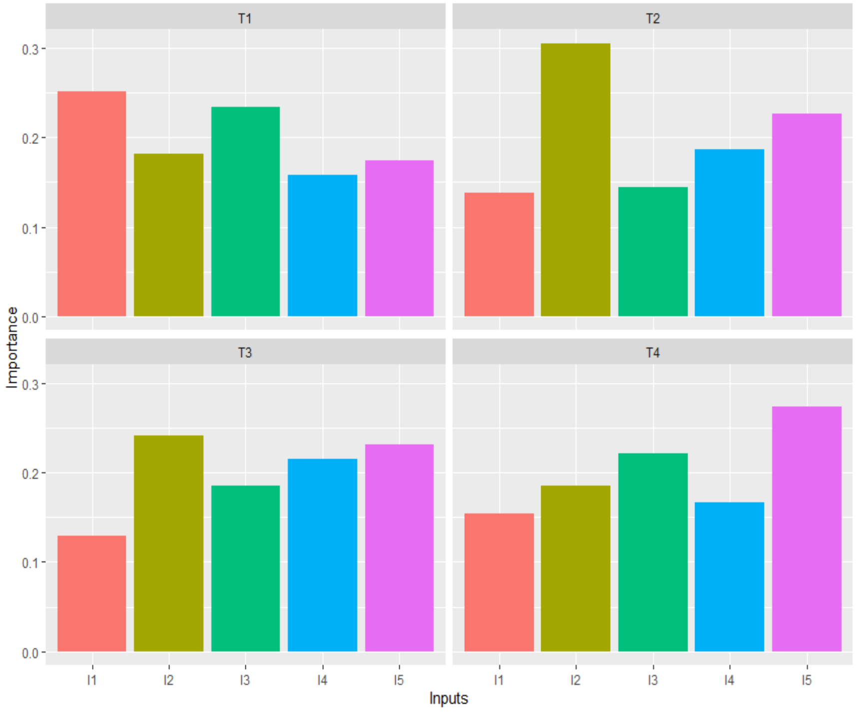


Figure 8-57. Variable importance of the 4 independent ANNs using Garson's.

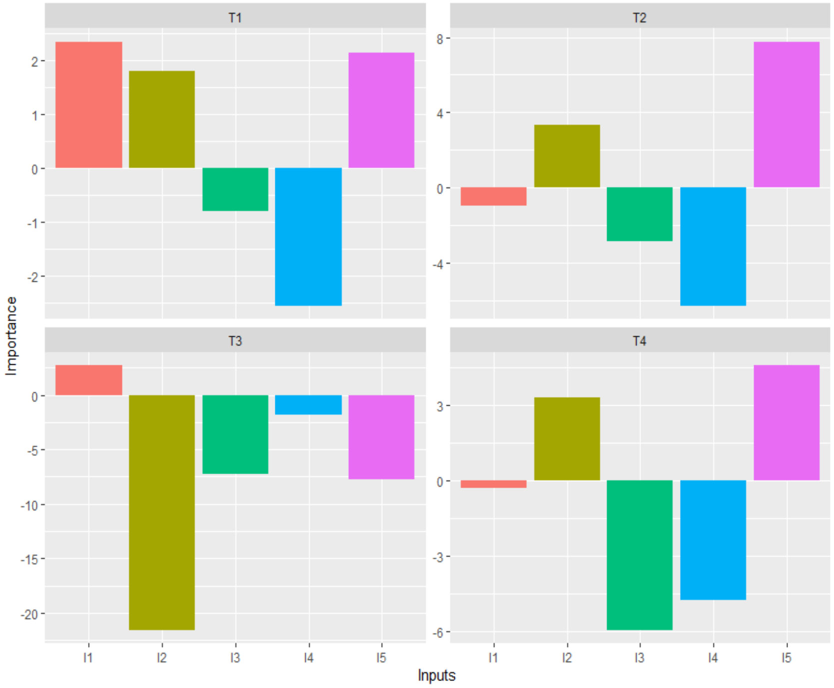


Figure 8-58. Variable importance of the 4 independent ANNs using Olden's algorithm.

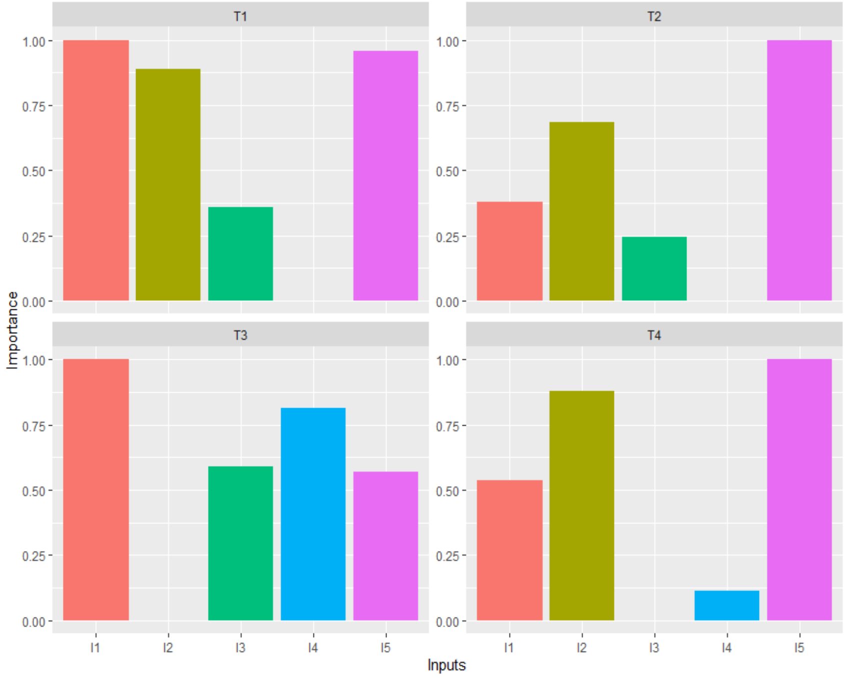


Figure 8-59. Variable importance of the 4 independent ANNs using connectivity map.

The main conclusion from **Figure 8-57** through **Figure 8-59** is that different importance analyses approaches did not provide a consistent result. One potential issue could be that ANN models are not suited for this data set when previous production rates are not used as predictors. One thing we observed is that when we re-trained ANN models, the importance analysis results would change dramatically. This may indicate that the trained ANN models had not converged. On the other hand, the importance analysis results in Subsection 8.3.2.1 for ANN models with previous production rates as predictors provide a similar result, especially for connection maps, to other methods such as the sensitivity analysis in Subsection 8.3.2 and Pearson's correlation coefficients in Subsection 8.3.4. Among the three importance analysis approaches, the connectivity map is recommended, because it provides a more consistent result than the other methods, it considers the impact of negative weights, and it is not strongly affected by extreme values as is Olden's connection weights algorithm.

8.4.2 Gaussian Processing Models

We also tried fitting GP models without using production rates at previous dates as predictors. The result (**Figure 8-60**) is slightly poorer than the ANN (**Figure 8-56**), especially for production well 1. Even though the predicted values have a relatively large prediction error, the GP model seems to be able to capture some effects from the injected water as we can see a similar up and down trend from the observed data and predicted values. However, for production wells 2 and 3, the fitted models are not accurate at all.

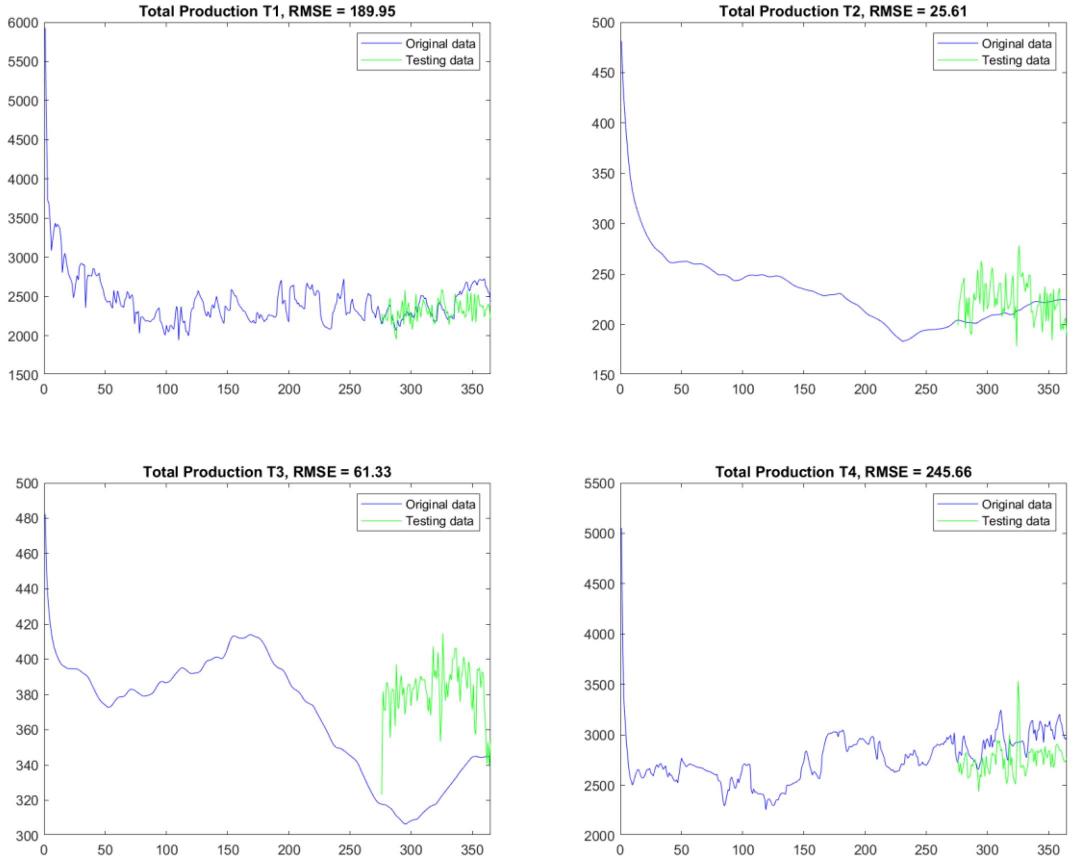


Figure 8-60. Predictions from a GP model without using total production rates as previous dates as predictors.

8.4.3 Linear Regression with Constraints

One way to improve the explainability of fitted models is to train a model that incorporate rules that govern reservoir physics. One idea is to assume that for all production wells the percentage of the sum of water received from each injection well should be equal to 1. For example, if 1 bbl water was injected into injection well 1, then a case that fits this assumption is if 0.7 bbl went to production well 1, and 0.1 bbl went to production wells 2~4, respectively. **Table 8-18** presents the coefficients of the fitted linear model with this constraint incorporated. Intercepts were also added into the model to help align the prediction results for overall mean production rates. The prediction accuracy of this model is illustrated in **Figure 8-61**.

Table 8-18. Coefficients of the fitted linear model with constraints.

Production/Injection	Intercept	Well 1	Well 2	Well 3	Well 4	Well 5
Well 1	497.10	0.92	0.44	0.05	0.21	0.41
Well 2	-891.17	0	0.03	0	0.18	0.09
Well 3	-815.83	0	0.07	0	0.23	0.12
Well 4	667.54	0.08	0.47	0.95	0.38	0.38

From **Table 8-18**, 92% of water injected into injection well 1 goes to production well 1 and 95% of water injected into injection well 3 goes to production well 4. Water in injection wells 2 and 5 equally goes to production well 1 and 4, while water in injection well 4 goes to all four production wells. The above findings are consistent with what was observed in the sensitivity analysis in Subsection 8.3.2. For example, if we removed the previous production rate, P(t-1), in **Table 8-16**, the most impactful input for production well 1 is the injection well 1. Similarly, the most impactful input for production well 4 is the injection from well 3.

As shown in **Figure 8-61**, prediction accuracy of the linear regression model with constraints is reasonable for production wells 1 and 4. It is better than the ANN and has a similar prediction accuracy to the GP models. However, there are still a couple of issues in the prediction. For example, the fitted linear model has poor fitted values in the early time points, although that is expected because they are potential outliers (see Section 8.2). Furthermore, the regression model with constraints cannot accurately predict production in wells 2 and 3, which is largely expected as both of those wells seem to be independent to the injected water.

We investigated three types of regression models: regression models without constraints, regression models with constraints but without intercepts, and regression models with constraints and with intercepts. We found that the intercept of the regression plays an important role to capture the overall mean of productions. RMSEs from those models can be found in **Table 8-20** at Subsection 8.4.

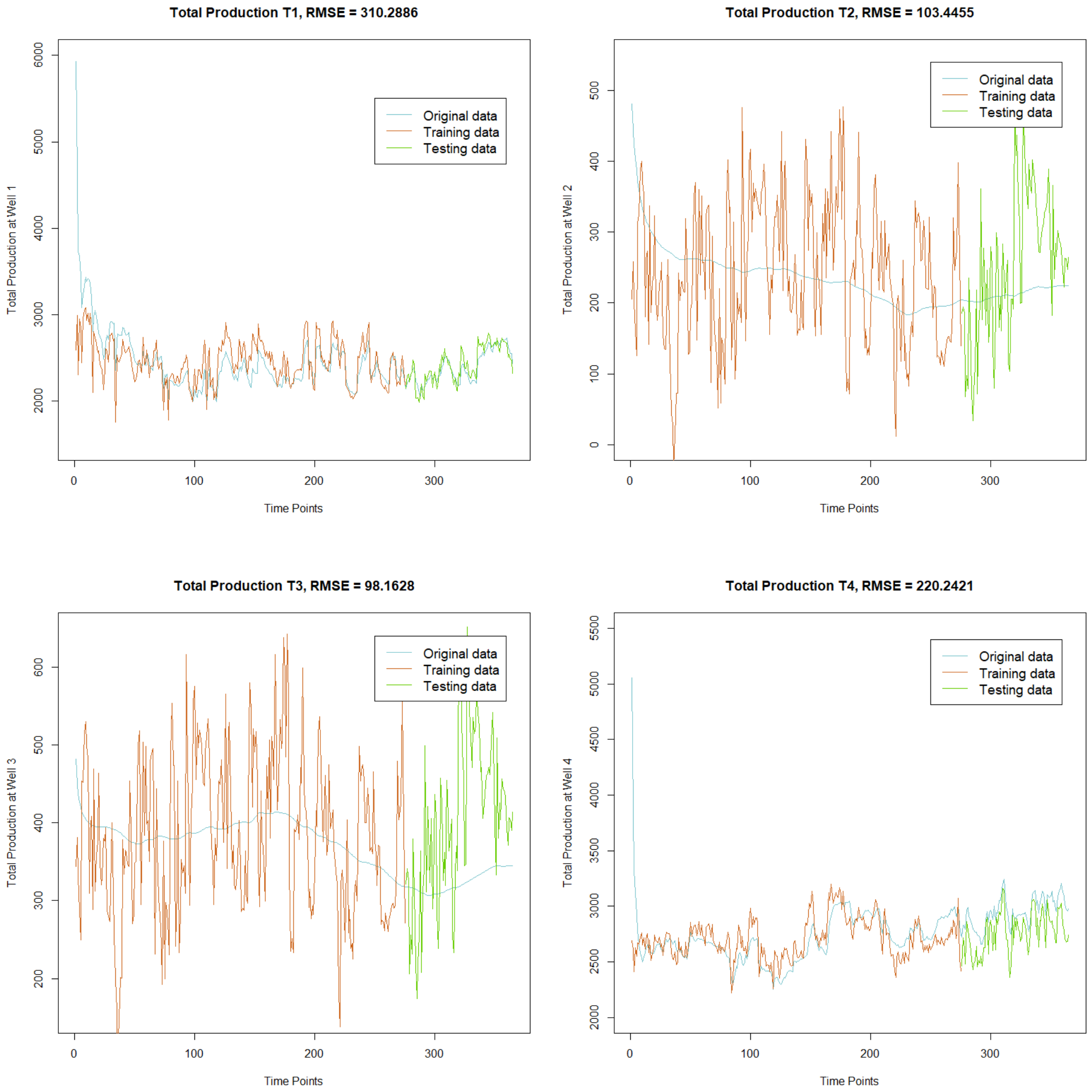


Figure 8-61. Predictions from a linear regression model with constraints without using total production rates as previous dates as predictors.

8.5 Results

Section 8.3 and Section 8.4 provided a brief introduction of several data-driven models. The goal of this section is to compare and summarize the prediction performance of all those data-driven models.

As mentioned in Section 8.3, the first 275 data points were treated as the training data and the remaining data points were treated as testing data for a fair comparison. We further used the same setting of those hyper-parameters in models trying to reduce the effect of parameters in different models. For example, we set the # of epochs = 1,000, batch size = 1 (for single output model) or batch size = 4 (for multiple outputs model) across all trained ANN models. Since there is some randomness in model training, we trained each model 10 times and took the mean values of the RMSE.

8.5.1 Time Series Forecasting

Table 8-19 shows the RMSE values from the models introduced in Section 8.3 with different window size (WS) options. Note that "-" denotes missing results, either because the analysis was not performed or cannot be performed (e.g., GP models interpolate data so it always provides perfect predictions on training dataset).

The RMSE shown in **Table 8-19** may not be the optimal results for all models, because there are several tuning parameters (e.g., number of epochs, batch size, loss function, optimizer, etc.) that in theory could be adjusted, but were fixed across all models for this study. In general, the multi-output ANN model with window size 3 seems to be the optimal choice, even though it does not provide the best prediction accuracy across all outputs. Particularly, the model has the best performance in predicting total production rates at wells 1 and 4, which may indicate that the ANN did capture the effect from previous total production rates and injection rates.

Table 8-19. RMSEs for testing and training sets of several data-driven time series forecasting models.

RMSE		Testing Set								Training Set							
Models	WS	T1	T2	T3	T4	WR1	WR2	WR3	WR4	T1	T2	T3	T4	WR1	WR2	WR3	WR4
Uni.LSTM	1	34	2.5	2.6	67	.10	.01	.006	.01	72	5.3	3.4	64	.012	.02	.017	.02
Uni.LSTM	3	32	5.7	2.8	81	.17	.01	.01	.015	45	3.6	2.6	72	.018	.05	.04	.03
Mul.LSTM	1	65	1.7	2.5	80	.02	.03	.02	.05	46	3.0	1.9	30	.01	.01	.004	.01
Mul.LSTM	3	56	3.4	1.9	59	.1	.01	.006	.02	45	3.6	2.6	30	.03	.05	.05	.02
Mul.ANN	1	33	1.8	2.1	63	.08	.009	.003	.009	83	2.3	1.8	39	.013	.01	.005	.02
Mul.ANN	3	15.5	2.7	1.2	16	.1	.01	.008	.01	29	3	2	28	.016	.01	.007	.02
Regression	1	37	0.9	0.9	88	.01	.01	.004	.005	60	2.6	2.2	34	.01	.001	.0004	.01
GP	1	73	2.1	14	166	-	-	-	-	-	-	-	-	-	-	-	-
GP	3	21	25	235	131	-	-	-	-	-	-	-	-	-	-	-	-

8.5.2 Regression Forecasting

Table 8-20 shows the RMSE values of the data-driven models introduced in Section 8.4. Note that to estimate the coefficients of the regression with constraints we used an optimization algorithm using the least-squares method. However, the optimization algorithm did not work well for water ratios. For example, all of the estimated coefficients were 0.25 which are the initial values in the optimization algorithm; this indicates that the optimization algorithm failed to produce a reasonable result. Thus, in **Table 8-20** only results for total production rates are shown.

Table 8-20. RMSEs for testing and training sets of several data-driven regression forecasting models.

RMSE	Testing Set				Training Set			
Models	T1	T2	T3	T4	T1	T2	T3	T4
Regression without constraints	82	25	23	219	294	40	62	198
Regression with constraints without intercepts	210	169	206	144	354	153	151	281
Regression with constraints	88	102	123	207	310	103	98	220
GP	211	26	61	246	-	-	-	-
ANN	89	38	72	211	516	73	44	348

8.6 Discussion

This report investigated several data-driven models for charactering reservoir connectivity and forecasting waterflood production from the viewpoints of time series forecasting and regression forecasting. While time series forecasting provides a much better prediction accuracy, regression forecasting typically has a better interpretability in the connectivity between injection and production. The recommended model for time series forecasting is the ANN not only because ANNs in general provide the best prediction accuracy but because this ANN has a relatively simple structure so it is possible to open its black-box neural networks to understand the connectivity between injection and production wells. For regression forecasting, the recommended model is the regression model with constraints. Even though the regression model without constraints has a slightly better performance in prediction accuracy, the regression model with constraints has a much better interpretability; specifically, its coefficients can be treated as the proportion contributed from each injection well to that production well.

The prediction performance of those trained models in this report seem to be not competitive compared to the results shown in other literature sources that used data-driven model analyzing the interwell connectivity. There are two potential reasons. First, the size of data in this experiment is relatively small (e.g., this experiment only has 365 observations, while [9] has 3,000 and [11] had around 4,500 observations, respectively). Thus, the lack of data would cause a poor model fitting especially for those complex deep learning models. Second, the goal of this experiment is to use historical data for building models to predict the future data, while most papers focused on the history-matching results. For example, [11] randomly selected 80% of the historical production and injection data as training set and the other 20% of the history data were used as testing set, which is an *interpolation* task. However, this report used the observations at the first 275 time points as training data to predict the remaining observations, which is a more difficult *extrapolation* task.

8.7 References

1. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

2. J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, “Design and Analysis of Computer Experiments,” Stat. Sci., vol. 4, no. 4, pp. 409–423, 1989.

3. A. Mittal, “Understanding RNN and LSTM,” Medium, Oct. 12, 2019. https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e (accessed Feb. 13, 2020).

4. Santner, T. J. and Williams B. and Notz W., The Design and Analysis of Computer Experiments, Second Edition. Springer-Verlag, 2019.

5. T. E. Fricker, J. E. Oakley, and N. M. Urban, “Multivariate Gaussian Process Emulators With Nonseparable Covariance Structures,” Technometrics, Feb. 2013, Accessed: Feb. 13, 2020. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00401706.2012.715835.

6. Han, G., Svenson, J. Leatherman, E., Moon, H. Chen, P.-H., Santner, T. J., “MATLAB Parametric Empirical Kriging (MPErK),” The Ohio State University, 2015.

7. Sobol, I. M., “On sensitivity estimation for nonlinear mathematical models,” jour Matem. Mod., vol. 2, no. 1, 1990.

8. U. Demiryurek, F. Banaei-Kashani, C. Shahabi, and F. G. Wilkinson, “Neural-Network based Sensitivity Analysis for Injector-Producer Relationship Identification,” in Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands, 2008, doi: 10.2118/112124-MS.

9. E. Artun, “Characterizing Reservoir Connectivity and Forecasting Waterflood Performance Using Data-Driven and Reduced-Physics Models,” in SPE Western Regional Meeting, Anchorage, Alaska, USA, 2016, doi: 10.2118/180488-MS.

10. E. Artun, “Characterizing interwell connectivity in waterflooded reservoirs using data-driven and reduced-physics models: a comparative study,” Neural Comput. Appl., vol. 28, no. 7, pp. 1729–1743, Jul. 2017, doi: 10.1007/s00521-015-2152-0.

11. H. Cheng, X. Han, P. Zeng, H. Yu, E. Osipov, and V. Vyatkin, “ANN based Interwell Connectivity Analysis in Cyber-Physical Petroleum Systems,” in 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), Helsinki, Finland, Jul. 2019, pp. 199–205, doi: 10.1109/INDIN41052.2019.8972285.

12. J. D. Olden, M. K. Joy, and R. G. Death, “An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data,” Ecol. Model., vol. 178, no. 3–4, pp. 389–397, Nov. 2004, doi: 10.1016/j.ecolmodel.2004.03.013.

13. B. Alizadeh, S. Najjari, and A. Kadkhodaie-Ilkhchi, “Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran,” Comput. Geosci., vol. 45, pp. 261–269, Aug. 2012, doi: 10.1016/j.cageo.2011.11.024.

14. Marcus W. Beck, “NeuralNetTools: Visualization and Analysis Tools for Neural Networks,” J. Stat. Softw., vol. 85, no. 11, pp. 1--20, 2018, doi: 10.18637/jss.v085.i11.

15. Garson, D.G., “Interpreting neural network connection weights,” AI Expert, vol. 6, pp. 47–51, 1991.

Chapter 9.

Evaluating Variable Importance in Time-Dependent Black Box Models: A Comparison of Strategies

Luan Lin, Jared Schuetter, and Srikanta Mishra

Chapter 9. Table of Contents

Chapter 9. List of Tables

Chapter 9. List of Figures

Chapter 9.

9.0 ABSTRACT

Measuring variable importance for computational models is an important task in many applications. It is always desirable to have a strategy that works for any model and could uncover the key predictors in the modeling. In this paper, we first review several commonly used variable importance strategies that are compatible with all machine learning or black box models and provide a comparative assessment of these strategies using an example from a subsurface geoscience application. Furthermore, we present a framework for making comparisons not only within but also between different time points for time-dependent models. We propose the relative importance score (*RIS*) and uncertainty importance factor (*UIF*), which allow users to intuitively interpret how variable importance changes over time.

9.1 Introduction

Developing statistical learning based “proxy” models using results from physics-based simulators for repetitive tasks such as uncertainty quantification or history matching is becoming commonplace in the petroleum geosciences [16]. These proxy models are typically developed for modeling univariate summary outputs or outputs at selected points in time using a combination of experimental design and response surface analyse [2, 6, 10, 18, 27]). In recent years, the attention has turned to the use of sampling-based designs coupled with more flexible data-driven modeling techniques such as neural networks, kriging (gaussian process) models, etc., to obtain better granularity in the results [1, 13, 12, 19, 21, 22, 23]. In many of these studies, the identification of key variables is typically done before building the proxy model with the use of screening techniques such as a 2-level Plackett-Burman design [28]. Thus, the issue of variable importance on the proxy model itself is generally not discussed. Furthermore, with time-dependent responses, it is demanding to gain insight into the dynamic characteristics of the system in addition to obtaining a computationally efficient and accurate proxy model. For example, the phenomena governing the system may change over time as critical conditions are met, and stakeholders might be interested in knowing whether and how the roles of different variables in the system change over the course of the simulation. It can be challenging to identify which variables are important at different time points and evaluate how those variables drive the model response over time.

Measuring variable importance for computational models is an important task in many applications. With a predictive model, it is also desirable to extract information about the relationships uncovered by the model. Researchers are often interested in knowing which predictors, if any, are important by assigning some type of importance scores to each variable, and this has resulted in variable importance techniques being developed independently in many disciplines [14, 25]. Some variable importance strategies are specific to a model’s parameterization or are linked to assumptions made by the models. However, researchers are more frequently encountering diverse sets of models, or ensembles of models, and because of this it is desirable to have a strategy that is more agnostic. That is, it will work for any model that can be characterized as a “black box” that converts a set of predictor values into one or more responses.

The objective of this paper is to describe several of the commonly used variable/feature importance strategies that are compatible with black box models and provide a comparative assessment of those strategies for an example subsurface geoscience application. A motivating factor in this regard is the development of ensemble modeling approaches that result in multiple acceptable models Schuetter, *et al.,* [24] thus requiring a set of model agnostic variable importance tools that can be compared across models. A second goal is to present a framework for making comparisons not only within but also between simulation time points for time-dependent models, allowing for a comprehensive understanding of how variable importance changes over time.

9.2 Methods

9.2.1 Variable Importance Strategies

Fundamentally, the idea of “importance” for a variable in a model is easy to understand. In the context of linear or linearized regression models, this is generally related to the fractional contribution to variance [17]. However, the concept is nebulous enough for black-box models that one can conceive of a variety of methods for quantifying it in a such models. As a result, there have been a number of different strategies developed for measuring variable importance, each of which is internally consistent with a set of assumptions, but amongst which there is a large degree of overlap [14]. There are also strategies that only apply to specific models, for example, the Gini-based importance for random forest models [4]. This section contains a description of some commonly used variable importance methods that are applicable to all machine learning or black box models, as well as an overview of the reasoning behind them. This is followed by an illustrative example to help the readers better understand each of the methods.

9.2.2 R^2 Loss

The first set of strategies considers variable importance from the perspective of model prediction performance. That is, an important variable is one that significantly impacts a model’s performance (i.e., explains much of the variability in the sampled data). In classical experiment design, the coefficients, their associated standard error, and the significance of each variable in an Analysis of Variance (ANOVA) model are often used to rank the predictor variables. While this approach assumes a linear model and is not directly extensible to black box models, one could generalize it by defining a measure for the quality of a model fit. One such example is the pseudo-R^2, given by

$$Pseudo\text{-}R^2 \equiv R^2 = 1 - \frac{SS_{model}}{SS_{error}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \qquad \textbf{(Eqn. 9-1)}$$

where y_i and $\widehat{y_i}$ are the true and black box predicted responses for the i^{th} observation, respectively, and $\bar{y}$ is the average response across the dataset. Conceptually, this pseudo-R^2 score measures how much better the model predictions are compared to a prediction of the mean response, with the maximum score being 1 and the minimum being, in theory, $-\infty$ although negative scores tend to be rare unless the models are quite poor.

To measure the impact of a predictor on the model, one can intentionally negate the relationship between the predictor and the response and see how that affects the pseudo-R^2. That is, one can compare the *loss in the R^2*, or “R^2 loss” from the original model that occurs when the predictor’s impact is negated.

To do this, there are two straightforward strategies. The first strategy, denoted as *Remove*, involves removing the predictor from the model, such that the original model uses the full set of predictors and the altered model uses all predictors except the one of interest. This is the same idea used in approaches like backward stepwise linear regression, where predictors are removed one at a time from the model to see if they explain a significant portion of the natural variation observed in the data.

The second strategy, denoted as *Permute*, involves permuting the values of the predictor across the observations in the dataset. In this case, the original model is the same as before, but the altered model now uses the randomized permuted predictor of interest in place of the original one. The random forest variable importance measure uses similar reasoning by permuting predictor values in out-of-bag samples and looking at changes in the mean squared error or Gini index.

In both strategies, the R^2 loss can be described as $R^2_{Loss} = R^2_{Original} - R^2_{Altered}$, where $R^2_{Original}$ is the R^2 obtained from the original black box model and $R^2_{Altered}$ is the R^2 calculated from a model trained with the predictor of interest either left out of the dataset (for the first strategy) or permuted across the observations (for the second strategy).

9.2.3 Partial Dependence (PDP) and Accumulated Local Effects (ALE) Plots

Partial Dependence Plots (denoted as *PDP),* [11, 9] were introduced with purpose of interpreting complex machine learning algorithms. Interpreting results from a linear regression model is straightforward because of the clear mapping between coefficients and the variables. However, interpreting machine learning and black box models (e.g., random forests, gradient boosting machines, and recently popular neural network models) is more difficult due to their complexity. *PDP*s use black-box model predictions to show the marginal effect of each variable on the predicted outcome, whether its effect is linear, monotonic, or more complex. The partial dependence function is estimated by calculating averages in the training data per the following expression:

$$\widehat{f_{X_S}}(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}\left(x_S, x_C^{(i)}\right), \qquad \textbf{(Eqn. 9-2)}$$

where X_S are the variables (usually one or two) for which the partial dependence function should be plotted, X_C are the other variables in the model $\hat{f}$, x_S are the values of the variables of interest, $x_C^{(i)}$ are the actual values of the rest of the variables in the dataset, and n is the number of instances in the dataset. Essentially, the model is holding the variables of interest fixed at each value they take in the dataset and averaging over the empirical distributions of the other variables. For large data sets, the grid size might be reduced by taking specific quantiles for each variable instead of using all unique values . We might expect that *PDP*s are “flat” for less important variables since less important variables tend to have little influence on the outcomes, while the variables whose *PDP*s vary over a wider range of the response are more likely to be important.

Similarly, plots of Accumulated Local Effects (denoted as *ALE*) [1] describe how variables influence the prediction of a machine learning model on average. If variables of a machine learning model are highly correlated, the *PDP*s are not reliable and cannot be trusted. For example, the computation of a *PDP* for a variable that is highly correlated with other variables might involve predictions of unlikely instances. As suggested by its name, *ALE* plots try to understand the local behavior of the response within small windows of each variable's support. *ALE* plots are faster and unbiased alternative to partial dependence plots. *ALE* plots average the changes in the predictions within grid cells evenly spaced over the range of the variable of interest. The value of the *ALE* can be interpreted as the main effect of the variable at a certain value compared to the average prediction of the data. For example, an *ALE* estimate of -2 at $x = 3$ means that when the variable has value 3, then the prediction is 2 less than the average prediction.

The variable importance score based on *PDP*s and *ALE* can be any measure of “flatness” of the partial dependence function and *ALE* plot function. An effective measure to use is the sample standard deviation for continuous variables and the range statistic divided by four for categorical variables [11]) The range divided by four provides an estimate of the standard deviation for small to moderate sample sizes.

9.2.4 Local Interpretable Model-Agnostic Explanations (LIME)

The Local Interpretable Model-Agnostic Explanations strategy, denoted as *LIME*, [19] uses surrogate interpretable models to explain individual predictions of machine learning or black box models. Although such models are now ubiquitous, it is nearly impossible to understand their inner workings, which raises the question of how much faith one should put in these models and their predictions. Thus, interpretable surrogate models are trained to approximate the predictions of underlying machine learning or black box models. Instead of training a global surrogate model, *LIME* aims to train local surrogate models to explain individual predictions. *LIME* attempts to understand the model by perturbing the values of a variable across the dataset and interpreting how the predictions change. Variable weights can then be extracted from a simple local model on the permuted dataset to explain local behavior. The procedures for training local surrogate models are:

- Select an observation of interest in the dataset.
- Perturb the dataset and generate the black box predictions for these new points.
- Weight the new samples according to their proximity to the observation of interest.
- Explain the prediction by interpreting the local model.

For example, a linear regression or LASSO model could be chosen as an interpretable surrogate model to help interpret the local predictions. The global variable importance score is then estimated by averaging the weights of the variable in the explanations across all instances.

The R package *lime* is used for local surrogate model interpretations Riberio, e*t al.*, [20]), where instances are sampled around x' by drawing nonzero elements of x' uniformly at random and the number of such draws is also uniformly sampled. Sparse linear models are used as explanations where K features are selected with LASSO and weights were learned via least squares. The results below use this package.

9.2.5 Shapley Additive Explanations (SHAP)

Shapley Additive Explanations [15] (denoted as *SHAP*) is a method to explain individual predictions based on the game-theoretically optimal Shapley values. The intuition of *SHAP* is similar with *LIME*, i.e. aiming to interpret and explain individual predictions. The goal of *SHAP* is to explain the prediction of an observation *x* by computing the contribution of each variable to the prediction. The *SHAP* explanation method is motivated by coalitional game theory, where the variable values of a data instance act as players in a coalition. Shapley values indicate how to fairly distribute the “payout” (i.e. the prediction) among the variables. The global importance score is then estimated by averaging the absolute Shapley values for each variable across all instances.

In game theory, the Shapley value is defined via a value function val of players in S. The Shapley value of a feature (i.e. variable) value is its contribution to the payout, weighted and summed over all possible feature value combinations:

$$\phi_j(val) = \sum_{S \subseteq \{x_1,\cdots,x_p\} \setminus \{x_j\}} \frac{|S|!\,(p-|S|-1)!}{p!} \left(val\left(S \cup \{x_j\}\right) - val(S)\right), \qquad \textbf{(Eqn. 9-3)}$$

where S is a subset of the features used in the model, x is the vector of feature values of the instance to be explained, and p is the number of features. Finally, $val_x(S)$ is the prediction for feature values in set S that are marginalized over features that are not included in set S:

$$val_x(S) = \int \hat{f}(x_1,\cdots,x_p)\, d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X)) \qquad \textbf{(Eqn. 9-4)}$$

The reasoning behind this approach is similar to the ideas behind R^2 loss and *PDP*s. Rather than using Monte Carlo sampling over the variables’ domains, as with *PDP*s, the *SHAP* method instead looks at all possible subsets of variables that either include or do not include the variable of interest x_j. The quantity $val\left(S \cup \{x_j\}\right) - val(S)$ then captures an R^2 loss type of quantity averaged over the variables that are not in the subset.

9.3 Illustrative Example

In this section, a simple artificial example is presented to illustrate the concept of variable importance and help with the interpretation of different variable importance measures.

Suppose that the response variable y is based on three variables spanning the domain [0,1], i.e. $y = 5f_1(x_1) + 3f_2(x_2) + f_3(x_3)$, where f_1, f_2, and f_3 are defined as below:

$$f_1(x) = 2|\mathrm{x} - 0.5| \qquad \textbf{(Eqn. 9-5)}$$

$$f_2(x) = \frac{\exp(x+1) - \exp\,(1)}{\exp(2) - \exp\,(1)} \qquad \textbf{(Eqn. 9-6)}$$

$f_3(x) \sim \mathrm{N}\left(0, \frac{1}{3}\right)$, scaled to [0,1] over the n random draws. **(Eqn. 9-7)**

We randomly sample $n = 100$ values from [0,1] to represent x_1, x_2, and x_3, then apply the equations above to generate the response y, pretending that is the observed dataset. The functions and observed dataset are presented in Figure 9-62. Based on the coefficients for different functions in definition of y, a reasonable guess might be that x_1 is the most important variable in prediction of y since function f_1 carries the greatest coefficient; and the next important variable would be x_2. Ideally, the variable importance measures will share this result.

We use an 80/20 Train/Test split and gradient boosting model (GBM; 100 trees) as the black-box model for illustration. In Figure 9-62, the blue and red dots represent the train and test data, respectively. R^2 loss for the three variables based on *Remove* and *Permute* strategies are presented in Table 9-21. For both strategies, x_1 and x_2 have a significantly large R^2 loss indicating both are important variables, while x_3 has a slight increase in R^2 (i.e., a negative R^2 loss), which shows that x_3 might play a noise role in the prediction and demonstrates the motivation in function design for f_3.

Figure 9-63 and Figure 9-64 show the calculated *PDP* and *ALE* for different variables using the illustrative example, respectively. These plots show the marginal effect of the three variables x_1, x_2, and x_3 on the prediction of y. The hash marks at the bottom show where the actual data fall. The standard deviation of *PDP* for x_1, x_2, and x_3 (as shown in Table 9-21) are 1.175, 0.697, and 0.053, respectively, which again supports the fact that x_1 and x_2 are the most and second most important variables, while x_3 is not important as a predictor. Similar patterns as *PDP*s are observed and produce the same conclusion.

Figure 9-65 and Figure 9-66 provide results of *LIME* and *SHAP* for the illustrative example. In Figure 9-65, data are presented using a heatmap, where blue indicates feature condition met while red indicates otherwise. The average weights from *LIME* and the average absolute Shapley value from *SHAP* are plotted. The average weights for x_1, x_2, and x_3 are 1.140, 0.652, and 0.076, respectively, while the average absolute Shapley values for x_1, x_2, and x_3 are 1.014, 0.646, and 0.063, respectively.

A summary of various variable importance measures is provided in Table 9-21. As shown in Table 9-21, the conclusions on variable importance are consistent for all the six strategies, i.e. x_1 is the most important variable, while x_3 is the least important variable.

Table 9-21. Summary of various variable importance measures for illustrative example.

Variable	**R^2 Loss (Remove)**	**R^2 Loss (Permute)**	**PDP**	**ALE**	**LIME**	**SHAP**
x_1	0.580	0.665	1.175	1.183	1.140	1.014
x_2	0.365	0.499	0.697	0.756	0.652	0.646
x_3	-0.008	-0.006	0.053	0.070	0.076	0.063

9.4 Measuring Variable Importance Across Model Types and Time Points

Another important goal of this paper is to investigate the dynamic characteristics of time-dependent systems with data from multiple time points. To identify different “driver” variables at different time points, we propose a unified framework that is applicable to all machine learning or black box models.

To measure variable importance across different time points, we first fit machine learning or black box models at each single time point. Next, in comparing the importance of different variables across time and for different strategies, the following concepts are proposed.

Raw score:

For a given strategy, let $R(t,p)$ be the **raw score** for predictor p at time t. Specifically, for R^2 loss strategies *Permute* and *Remove*,

$$R(t,p) = R_t^2(Full) - R_{p,t}^2(Modified), \qquad \textbf{(Eqn. 9-8)}$$

where $R_t^2(Full)$ is the pseudo-R^2 for the full model at time t and $R_{p,t}^2(Modified)$ is the pseudo-R^2 for the modified model at time t, where predictor p was either removed from the model or permuted across the dataset.

For strategies *PDP* and *ALE*,

$$R(t,p) = \begin{cases} \sum_{i=1}^{n} \left(g_{p,t}(i) - \bar{g}_{p,t}\right)^2 / n - 1 & \text{if predictor } p \text{ is continuous} \\ \left[\max_{i=1,\dots,n} g_{p,t}(i) - \min_{i=1,\dots,n} g_{p,t}(i)\right] / 4 & \text{if predictor } p \text{ is categorical} \end{cases} \qquad \textbf{(Eqn. 9-9)}$$

where $g_{p,t}(i)$ is the *PDP* or *ALE* response at time t for predictor p at the value given by the i^{th} observation.

For strategies *LIME* and *SHAP*,

$$R(t,p) = \frac{1}{n} \sum_{i=1}^{n} w_{t,p}(i), \qquad \textbf{(Eqn. 9-10)}$$

where $w_{t,p}(i)$ is the weight given to predictor p for the i^{th} observation using the *LIME* or *SHAP* methods, respectively.

Normalized score:

The **normalized score** is defined as

$$N(t,p) = \frac{R(t,p)}{C(t)}, \qquad \textbf{(Eqn. 9-11)}$$

where $C(t)$ is a strategy-specific normalization constant at time t. Note that for strategies *Permute* and *Remove*, $C(t) = R_t^2(Full)$. For strategies *PDP*, *ALE*, *LIME*, and *SHAP*, $C(t) = \bar{y}_t$, the average response at time t.

Relative Importance Score:

To compare the importance of variables within a time point, the **relative importance score (RIS)** may be used. It is defined as follows, where $N(t,p)$ is the normalized score defined earlier.

$$RIS(t,p) = \frac{N(t,p)}{\max_{p} N(t,p)} \qquad \textbf{(Eqn. 9-12)}$$

Uncertainty Importance Factor:

To compare the importance of variables across *all time points*, the **uncertainty importance factor (*UIF*)** may be used (Mishra, Deeds, & Ruskauff, 2009). It is given by

$$UIF(t,p) = \frac{N(t,p)}{\max_{t,p} N(t,p)}. \qquad \textbf{(Eqn. 9-13)}$$

By normalizing $N(t,p)$ across time and predictors, UIF provides a globally relative score with respect to both time points and predictors by bounding the normalized scores $N(t,p)$ at 1. Alternatively, by normalizing $N(t,p)$ across predictors at each time point, RIS provides a comparison of relative scores across predictors at each single time point by scaling the normalized scores $N(t,p)$ within that time point. The RIS could be interpreted as the relative dominance for the predictors at a certain time point – with $RIS = 1$ for the most important variable at that time point. In contrast, the magnitude of the UIF represents the relative effect size of the importance across predictors and time points. The magnitude of the UIF is comparable across different time points, while this is not the case for the RIS.

When considering only a single time point, the ranks of the predictors by UIF are mathematically equivalent to their ranks by RIS. Also, for a given time point the ratios of UIF for two different predictors p_1 and p_2 are equivalent to the ratios of the RIS, i.e.,

$$\frac{UIF(t,p_1)}{UIF(t,p_2)} = \frac{RIS(t,p_1)}{RIS(t,p_2)}. \qquad \textbf{(Eqn. 9-14)}$$

Thus, while different, the UIF and RIS are closely related. For example, it is possible that the UIF for a predictor p could be greater at time t_1 than at time t_2 (i.e., $UIF(p,t_1) > UIF(p,t_2)$), but the RIS at those times could show the reverse (i.e., $RIS(p,t_1) < RIS(p,t_2)$). In this case, although the effect size of p on predicted outcome is greater at time t_1 than t_2, its dominance is greater at t_2 than t_1; the effect sizes of other predictors might be even smaller which still makes p the most important.

9.5 An Example Application

9.5.1 Problem Description

The variable importance strategies discussed above are applied to black-box surrogate models developed from full-physics time-dependent simulations of CO_2 injection into a deep saline formation as described in detail elsewhere [23]. The system being studied represents a single-well injecting supercritical CO_2 into a bounded 2-D radial-cylindrical formation (storage reservoir) initially filled with brine. The model domain consists of a porous and permeable heterogeneous reservoir, overlain by a low-permeability cap rock. The top of the cap rock, the bottom of the reservoir and the lateral boundary are all assumed to be no-flow boundaries (Figure 9-67). The simulations are executed in the numerical simulator Generalized Equation of state Model GEM® developed by the Computer Modeling Group (CMG) (Computer Modelling Group LTD. 2014). GEM is a robust, multidimensional and fully compositional reservoir simulator that is widely used as one of the standard simulators to model the flow of three-phase, multicomponent fluids in the oil and gas industry, as well as for other subsurface energy resource applications.

Running a simulation requires the specification of the nine input parameters listed in Table 9-22 and results in a number of responses over a 30-year simulation period. Of these responses, the average reservoir pressure within the model domain was chosen as the metric of interest for the present analysis. In the original study, a number of classical experimental design techniques were used such as Box-Behnken (BB), augmented pairs (AP), and Plackett-Burman (PB) designs using a framework that involves a low (-1), reference (0), and high (+1) value. The present study uses maximin Latin hypercube sampling (MM) and maximum entropy (ME) designs in a sampling framework, where the values of the inputs for each run are sampled over the 9-dimensional unit hypercube $[0, 1]^9$ and then converted back to the original predictor scale using the distributions shown in the rightmost column of Table 9-22. The distributions include triangular, log-triangular, and equally likely discrete distributions over the range of the experimental design values. For triangular distributions, denoted $\mathrm{T}(a, b, c)$, and log-triangular distributions, denoted $\ln\mathrm{T}(a, b, c)$, the parameters a, b, and c represent lower limit, upper limit, and mode, respectively. Since the BB design for 9 input parameters has 97 unique runs, all of the designs were set at a size of 97 runs for the comparison. Thus, the training dataset was a combination of two space-filling designs, i.e. 97-run maximin Latin hypercube sample and 97-run maximum extropy design (194 runs total). In addition, performance of modeling was measured over an independent validation dataset, i.e. Latin hypercube sampling (LHS) simulation data. Each run had average reservoir pressure measured for a unique combination of the nine input parameters at 19 time points arranged in approximately logarithmic fashion from initialization through 30 years. Details on the various design options can be found in (Schuetter and Mishra [23]).

Table 9-22. CMG-GEM Inputs for Sampling Designs (MM, ME).

	Parameter	Experimental Design Values			Sampling Design Distributions
		Ref.	Low	High	
1	h_R	150	50	250	T(50, 250, 150)
2	h_{CR}	150	100	200	T(100, 200, 150)
3	$k_{avg,R}$	46	12	220	---
4	$k_{avg,CR}$	0.02	0.002	0.2	lnT(0.002, 0.2, 0.02)
5	k_V/k_H	0.1	0.01	1	lnT(0.01, 1, 0.1)
6	q	0.83	0.33	1.33	{0.33, 0.83, 1.33} Discrete with equal probability
7	ϕ_R	0.12	0.08	0.18	T(0.08, 0.18, 0.12)
8	ϕ_{CR}	0.07	0.05	0.1	T(0.05, 0.10, 0.07)
9	I_k	Random	Increase from top	Decrease from top	{“Random”, ”Inc”, ”Dec”} Discrete with equal probability

9.5.2 Model Fitting Approaches

In order to evaluate variable importance using the strategies outlined above, a specific machine learning or black box model needs to be specified. To illustrate and compare how these variable importance strategies work with different machine learning or black box models, the selection of models in Table 9-23 are used as surrogate models to implement the variable importance strategies discussed earlier.

Table 9-23. Model Fitting Approaches

Model	Notation	Description
Quadratic Model	Quadratic	Fit a quadratic polynomial model to the response. The quadratic polynomial model includes all linear, quadratic, and pair-wise cross-product terms between predictor variables.
Quadratic Model with LASSO Variable Selection	Quad LASSO	LASSO (Least Absolute Shrinkage and Selection Operator) regression is a technique to perform variable selection. It is done by adding a penalty term to the least squares term in the objective function for linear regression.
Kriging Model	Ord Kriging Univ Kriging	Kriging, or Gaussian Process Regression, is a method of interpolation for which the interpolated values are modeled by a Gaussian process that combines a trend term (typically a linear model) and a covariance structure for points that are close in space. Ordinary kriging (Krig) assumes a scalar trend, whereas universal kriging uses a parametric trend term. In this study, quadratic terms are included in the trend term for universal kriging (Krig2).
Multivariate Adaptive Regression Splines	MARS	MARS models (Friedman [8]) approximate the response surface using a collection of simple step and hinge functions.
Additivity and Variance Stabilization	AVAS	The AVAS model (Tibshirani [25] and Breiman and Friedman [5]) uses a non-parametric, iterative procedure to find some transformation of the responses that can be represented as a sum of transformed predictors.

Model	Notation	Description
Random Forest	RF	A random forest {Breiman [4]) is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees. This method works by averaging multiple decision trees, trained on different parts of the same training set, with the goal of reducing the variance when using a single tree while simultaneously improving accuracy.
Gradient Boosting Model	GBM	Gradient boosting models (Friedman [9] and Elith, *et al*. [7]) are similar in structure to random forests, but decision trees in the ensemble are trained sequentially. Each new tree focuses on improving predictions for observations that are not already being predicted well by the other trees trained in previous steps.

9.6 Results

These eight modeling approaches from Table 9-23 were fit at each of the 19 time points of interest. Figure 9-68 presents the pseudo-R^2 for each model approach along time points. Some model approaches worked well at early time points, but not at later time points, while some model approaches generally behaved worse than some others. The illustration of variable importance results was then based on “Krig2”. As for this example dataset, Universal Kriging (“Krig2”) consistently had high pseudo-R^2 along all 19 time points and was considered the optimal modeling approach for this specific dataset.

Variable importance results were then computed for the surrogate models fitted at each of the time points using the six different strategies, *Remove*, *Permute*, *PDP*, *ALE*, *LIME* and *SHAP*. For the average pressure response, it was found that importance scores and rankings change across time points, indicating different variables driving the average reservoir pressure throughout the course of the simulation . To illustrate this, Figure 9-69 and Figure 9-70 present the Relative Importance Scores for average pressure at time points 2 and 19, respectively. From these plots, one can observe that at early time points, reservoir thickness and cap rock thickness are the only two variables that are important. One exception is the AVAS model for the *Remove* strategy, which identified reservoir porosity as the only important variable. The lack of consistency in this case might due to the poor fit of AVAS model at early time points (see Figure 9-68). We generally recommend selecting modeling approaches that generally had good performance along all time points. When time passes, however, injection rate starts to play a more important role and later becomes the most important variable driving the response. **Error! Reference source not found.** through Figure 9-76 present the UIF and RIS values for “Krig2” for each of the nine input predictors across time. “Krig2” is selected because of it generally fits the data well across different time points. For both UIF and RIS curves, there is a clear decreasing trend for reservoir thickness and cap rock thickness and an increasing trend for injection rate over time. When looking for the dominating or important predictors at different time points, RIS might provide a more straightforward view by assigning a score of one for the most important predictor at each time point. However, the magnitude of the RIS can only be interpreted and compared within each time point. Alternatively, UIF allows for a comparison of the magnitude of the scores across different time points to evaluate the overall importance of these predictors. For example, the *SHAP* approach in Figure 9-75 assigns much more importance to the injection rate in late time points than to the reservoir and cap rock thickness in early time points; this is evident in the UIF, but not the RIS.

9.7 Concluding Remarks

This study provides a comprehensive evaluation of different variable importance strategies using a common model-agnostic framework. Variable importance scores were found to vary depending on the modeling methods and importance measurement strategies used, but the top variables are usually consistently identified as being important. Among the modeling methods under consideration in this study, most were fast to train except the Kriging model. Among the variable importance strategies, *Permute*, *PDP*, and *ALE* were the fastest to compute. The *Remove* method was time-consuming, since it required continuously refitting models when a variable was removed. The computation times of *LIME* and *SHAP* varied with the size of the dataset. Appropriate modeling methods should be selected depending on the specific prediction problem being considered. Based on the analysis described here, *Permute* is a good recommendation because of its efficiency and consistency. Furthermore, it is beneficial to select a model approach which generally fits the data well across different time points.

Guidance on using UIF or RIS as the single variable importance metric of choice depends on the ultimate goal of the variable importance comparison. If one is interested in evaluating which variables are important at a specific time point, the RIS is a more straightforward way to identify dominant variables. Alternatively, if one is interested in comparing the magnitude/effect size of the variable importance across different time points, UIF is the more appropriate measure.

Finally, we note that in this work, statistical models were utilized as a surrogate model, and variable importance scores were evaluated separately across different time points for comparison. A natural extension would be to develop a unified approach which integrates data from all time points for modeling and variable importance investigation. A more systematic and accurate fit is expected by integrating the time variable in the modeling. Meanwhile, we expect to develop corresponding unified time-based feature importance strategy.

The primary contributions of this paper can be summarized as follows:

- Evaluation of a set of popular model agnostic variable importance techniques for a common problem representative of subsurface geoscience

- Two new metrics (UIF and RIS) introduced to handle variable importance rankings across different points in time

9.8 References

1. Amini, S., & Mohaghegh, S. (2019). Application of Machine Learning and Artificial Intelligence in Proxy Modeling for Fluid Flow in Porous Media. *Fluids, 4*(3), 126.

2. Anbar, S. (2010). *Development of a predictive model for carbon dioxide sequestration in deep saline carbonate aquifers.* Paper Presented at the SPE Annual Technical Conference and Exhibition.

3. Apley, D. W. (2016). Visualizing the effects of predictor variables in black box supervised learning models. *arXiv preprint*, arXiv:1612.08468.

4. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32.

5. Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association, 80*(391), 580-598.

6. Ekeoma, E., & Appah, D. (2009). *Latin hypercube sampling (lhs) for gas reserves.* Paper Presented at the Nigeria Annual International Conference and Exhibition.

7. Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813.

8. Friedman, J. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 1-67.

9. Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189-1232.

10. Ghomian, Y., Sepehrnoori, K., & Pope, G. A. (2010). *Efficient investigation of uncertainties in flood design parameters for coupled CO2 sequestration and enhanced oil recovery.* Paper Presented at the SPE International Conference on CO2 Capture, Storage, and Utilization.

11. Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). A simple and effective model-based variable importance measure. *arXiv preprint*.

12. Guerillot, D., & Bruyelle, J. (2016). *History Matching Methodology Using an Optimal Neural Network Proxy and a Global Optimization Method.* In Third EAGE Integrated Reservoir Modelling Conference. doi:10.3997/2214-4609.201602403

13. Hamdi, H., Couckuyt, I., Sousa, M., & Dhaene, T. (2017). Gaussian Processes for history-matching: application to an unconventional gas reservoir. *Computational Geosciences, 21*, 267-287. doi:10.1007/s10596-016-9611-2

14. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning.* New York, NY: Springer.

15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.

16. Mishra, S., & Datta-Gupta, A. (2017). *Applied Statistical Modeling and Data Analytics: A Practical Guide for the Petroleum Geosciences.* New York, NY: Elsevier.

17. Mishra, S., Deeds, N., & Ruskauff, G. (2009). Review Paper - Global Sensitivity Analysis Techniques for Groundwater Models. *Ground Water, 47*(5), 730-747.

18. Osterloh, W. T. (2008). *Use of multiple-response optimization to assist reservoir simulation probabilistic forecasting and history matching.* Paper Presented at the SPE Annual Technical Conference and Exhibition.

19. Rana, S., Ertekin, T., & King, G. (2018). An Efficient Probabilisitc Assisted History Matching Tool Using Gaussian Processes Proxy Models: Appliation to Coalbed Methane Reservoir. *Society of Petroleum Engineers*. doi:10.2118/191655-MS

20. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM.

21. Rwechungura, R., M., D., & Kleppe, J. (2011). Advanced History Matching Techniques Reviewed. *Society of Petroleum Engineers*. doi:10.2118/142497-MS

22. Sayyafzadeh, M. (2017). Reducing the Computation Time of Well Placement Optimization Problems Using Self-Adaptive Metamodeling. *J. Petrol. Sci. Eng., 151*, 143-158.

23. Schuetter, J., & Mishra, S. (2015). *Experimental Design or Monte Carlo Simulation Strategies for Building Robust Surroate Models.* Houston, TX: Paper 174905, SPE Annual Technical Conference and Exhibition.

24. Schuetter, J., Mishra, S., Lin, L., & Chandramohan, D. (2019). *Ensemble Learning: A Robust Paradigm for Data-Driven Modeling in Unconventional Reservoirs.* Unconventional Resources Technology Conference. doi:10.15530/urtec-2019-929

25. Tibshirani, R. (1988). Estimating transformations for regression via additivity and variance stabilization. *Journal of the American Statistical Association, 83*(402), 394-405.

26. Wei, P., Lu, Z., & Jingwen Song. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety, 142*, 399-432.

27. Wriedt, J., Deo, M., Han, W. S., & Lepinski, J. (2014). A methodology for quantifying risk and likelihood of failure for carbon dioxide injection into deep saline reservoirs. *International Journal of Greenhouse Gas Control, 20*, 196-211.

28. Yin, J., Park, H., Datta-Gupta, A., & King, M. (2011). A Hierarchical Streamline Assisted History Matching Approach with Global and Local Parameter Updates. *J. Pet. Sci. Eng., 80*(1), 116-130.

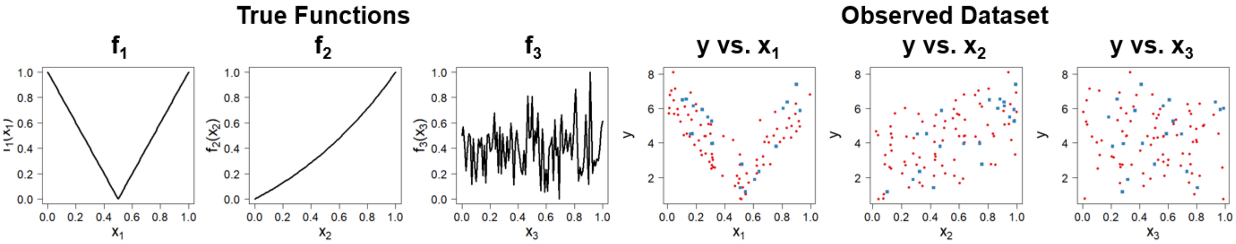


Figure 9-62. True functions and observed dataset for illustrative example.

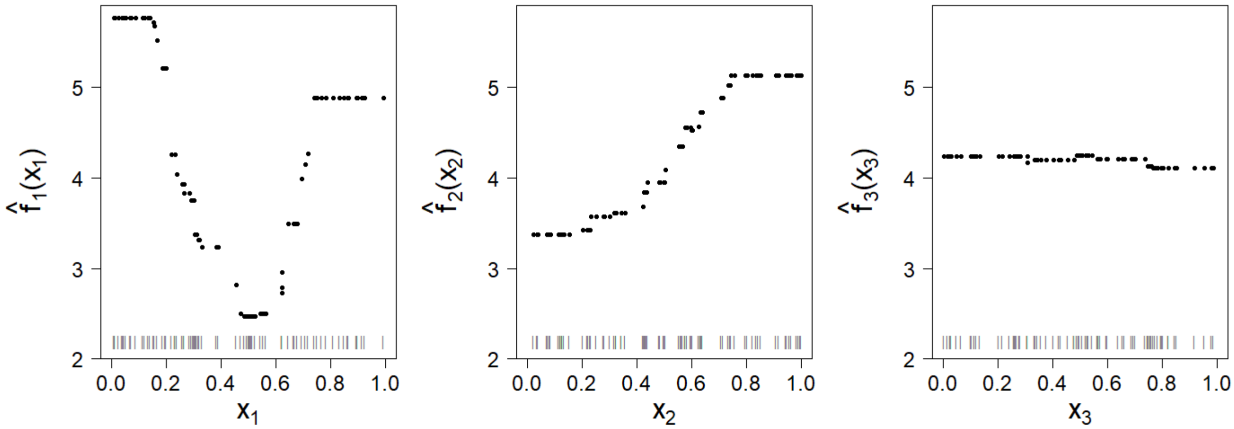


Figure 9-63. Partial Dependence Plots (PDPs) for illustrative example.

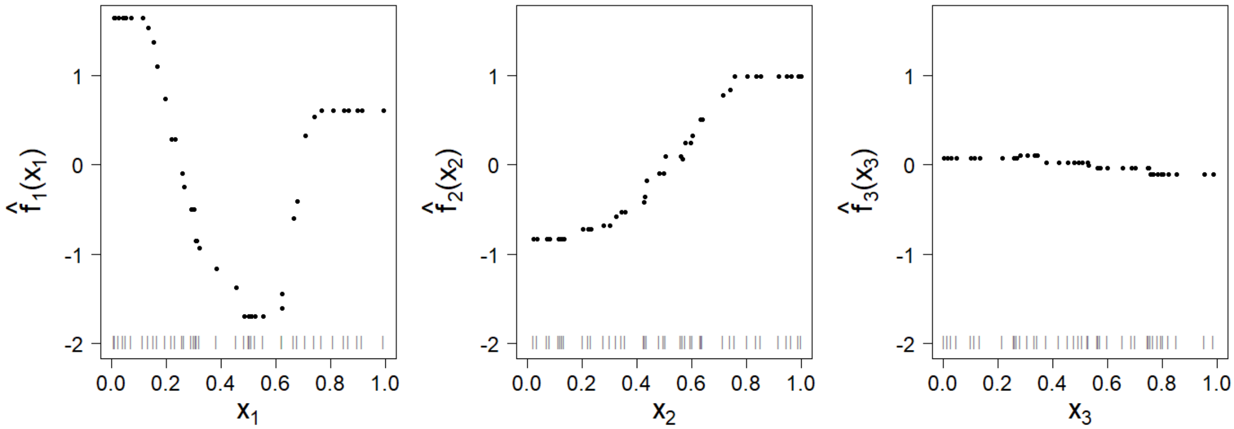


Figure 9-64. Accumulated Local Effects (ALE) for illustrative example.

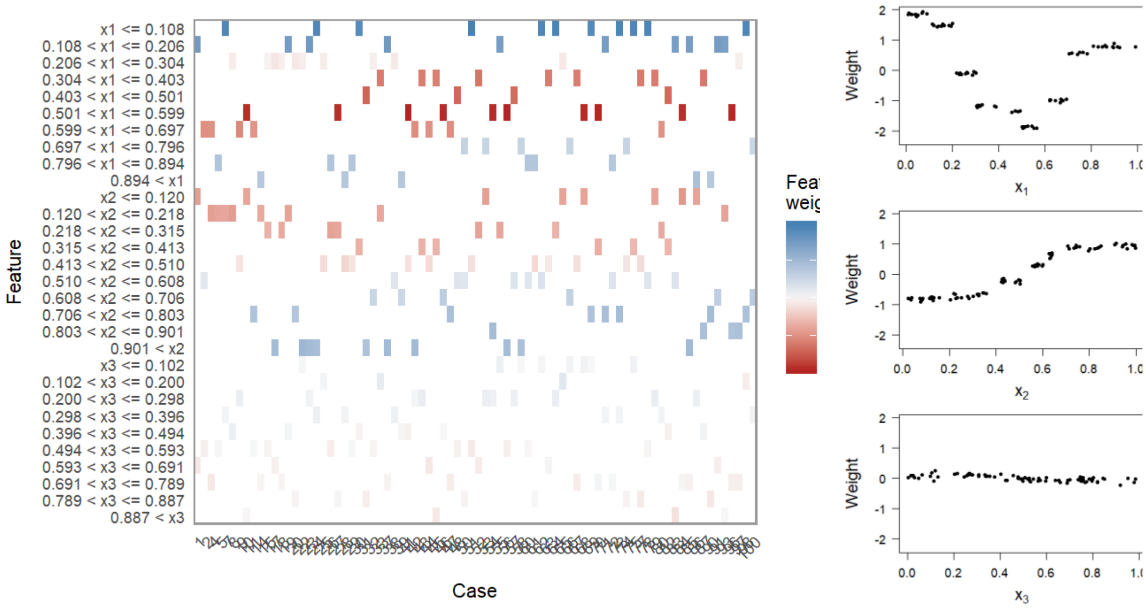


Figure 9-65. Local Interpretable Model-Agnostic Explanations (LIME) for illustrative example.

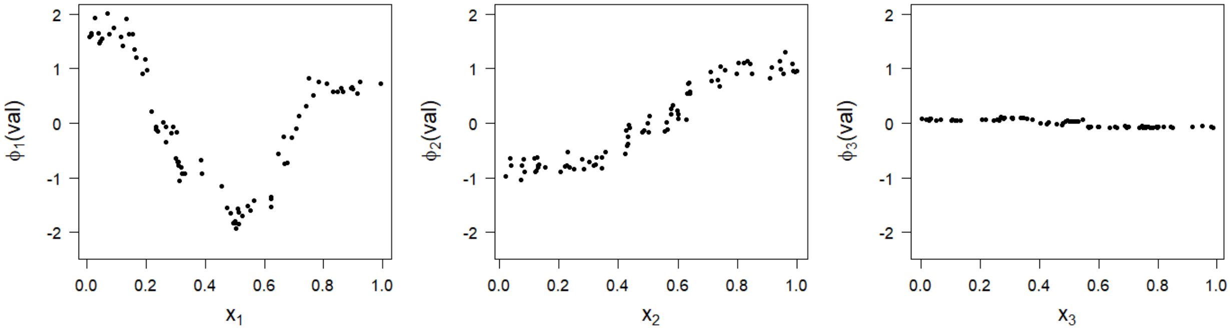


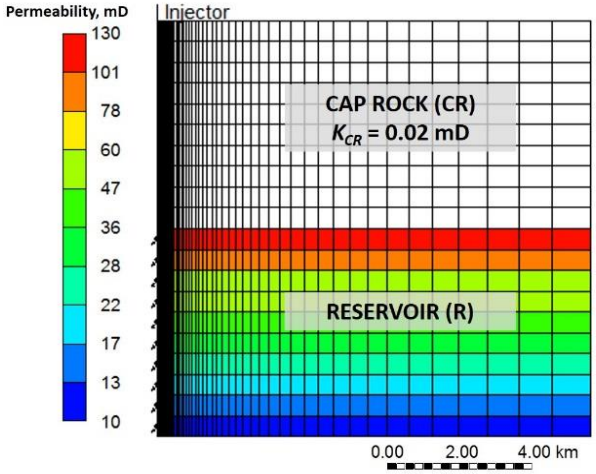
Figure 9-66. Shapley Additive exPlanations for illustrative example.

Inputs

Thickness of reservoir (h_R)
Thickness of caprock (h_{CR})
Average horizontal permeability of reservoir ($k_{avg,R}$)
Average horizontal permeability of caprock ($k_{avg,CR}$)
Anisotropy ratio (k_V/k_H)
CO_2 Injection rate (Q)
Porosity of reservoir (ϕ_R)
Porosity of caprock (ϕ_{CR})
Capillary pressure model of caprock ($P_{C,CR}$)
Relative permeability model of reservoir ($k_{r,R}$)
Indicator for permeability layering (I_k)

Responses

Total Storage Efficiency
Maximum Plume Radius
Average Pressure



Inputs in gray were fixed, while the ones in blue were varied in the space-filling design. The response of interest is in blue, as well.

Figure 9-67. Model geometry and variables of interest.

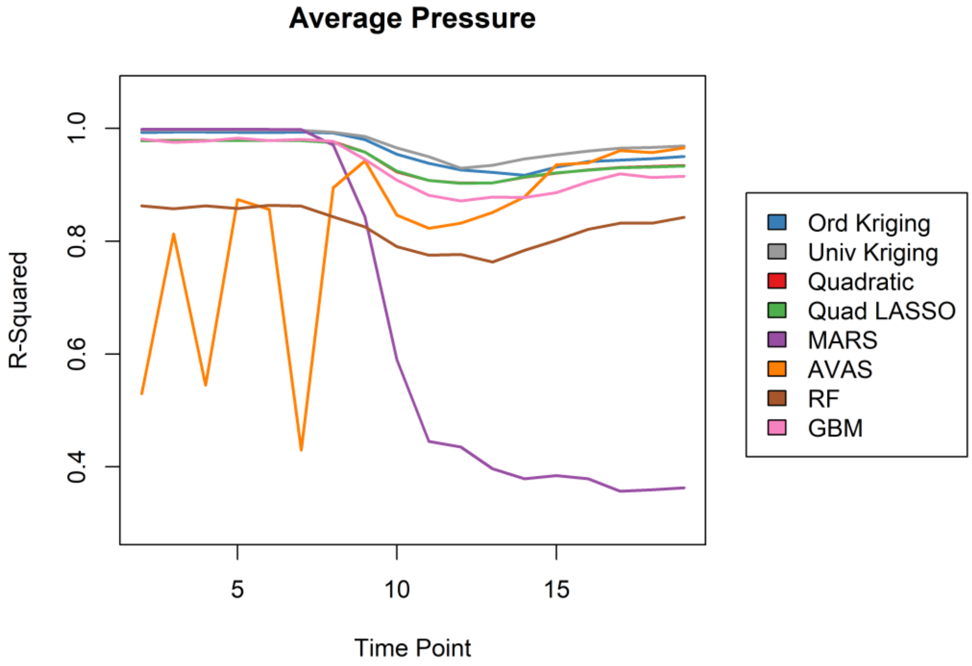


Figure 9-68. Pseudo-R2 scores for different models across time.

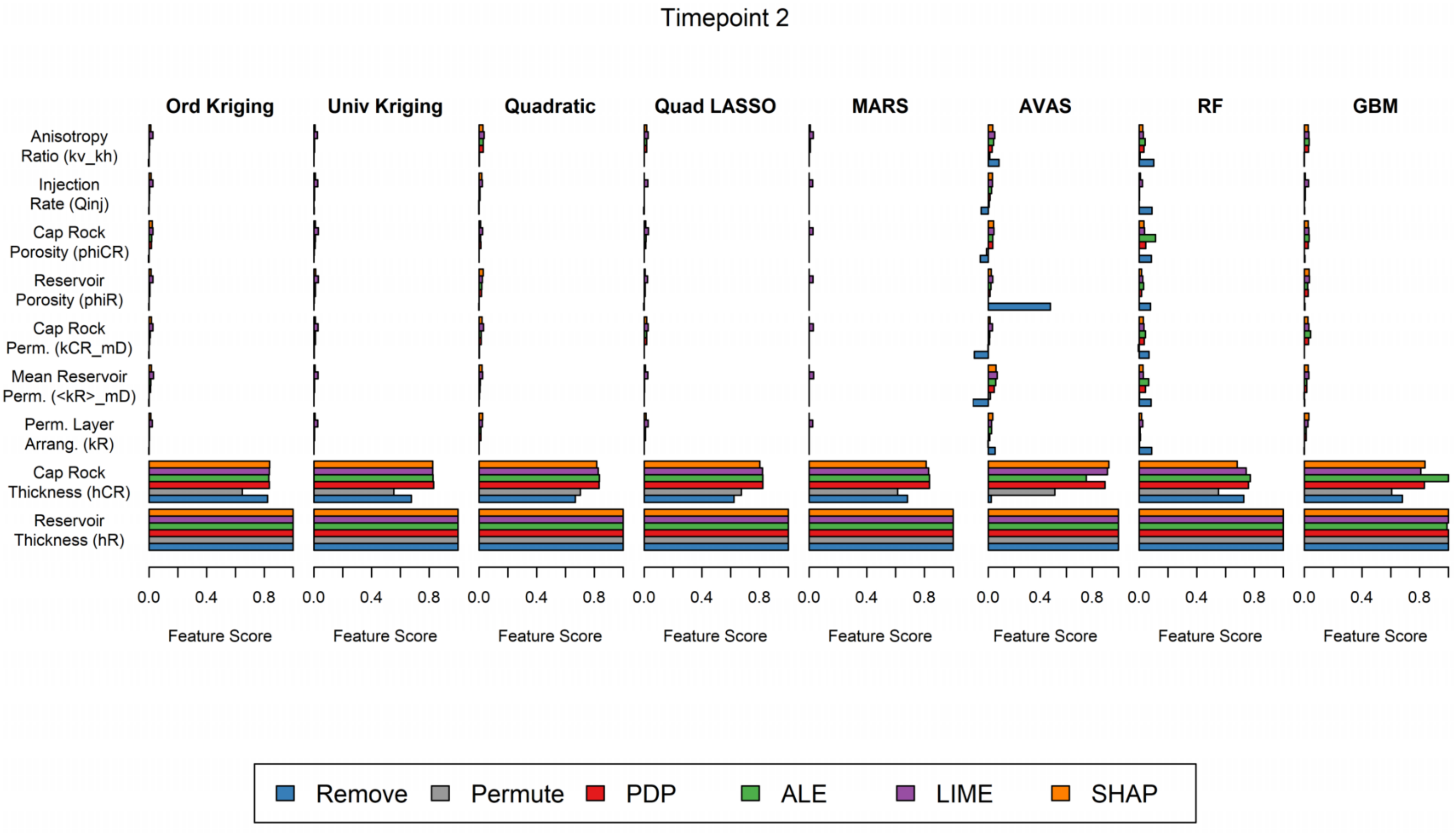


Figure 9-69. Relative Importance Scores for Average Pressure at Time Point 2.

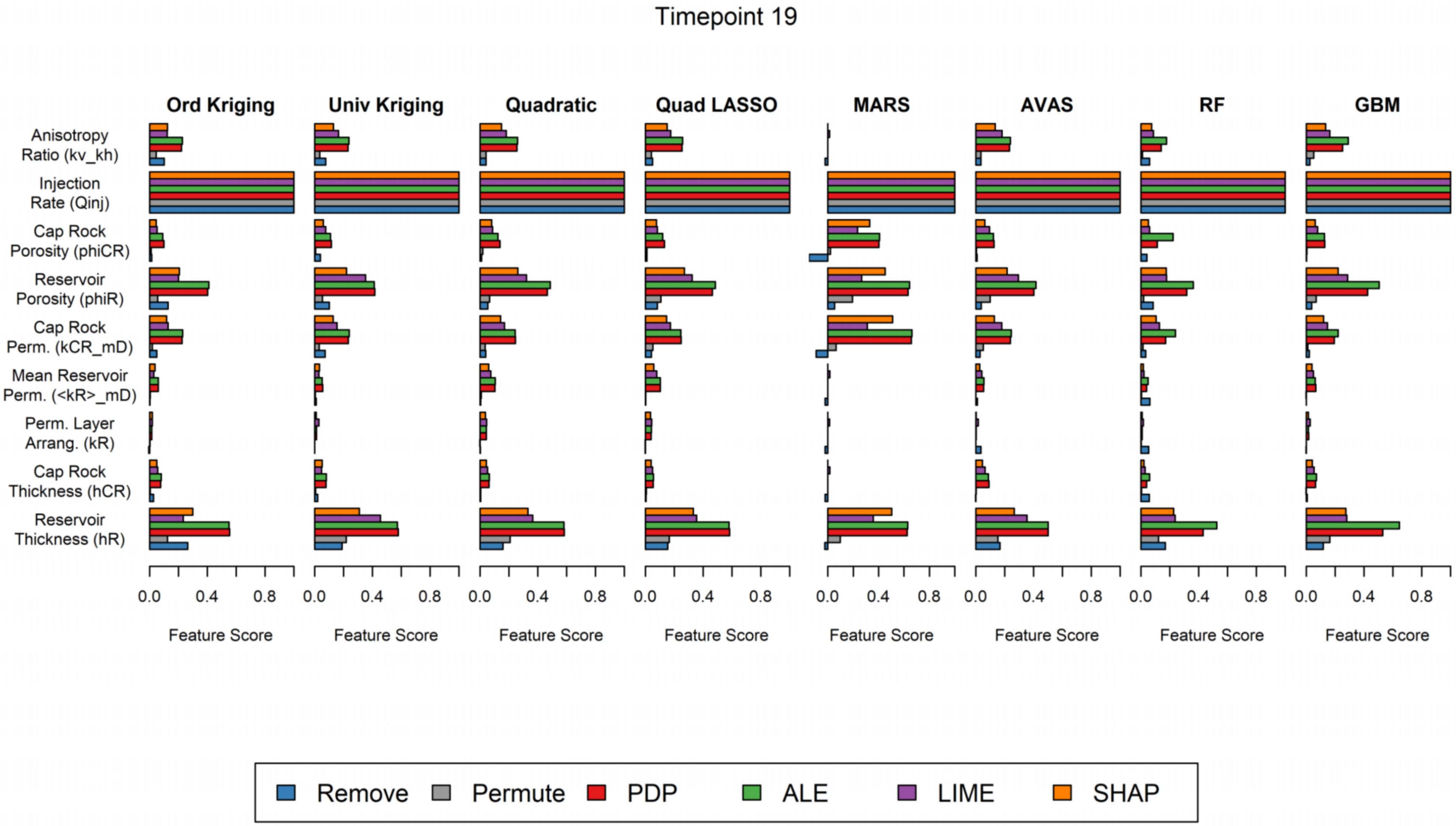


Figure 9-70. Relative Importance Scores for Average Pressure at Time Point 19.

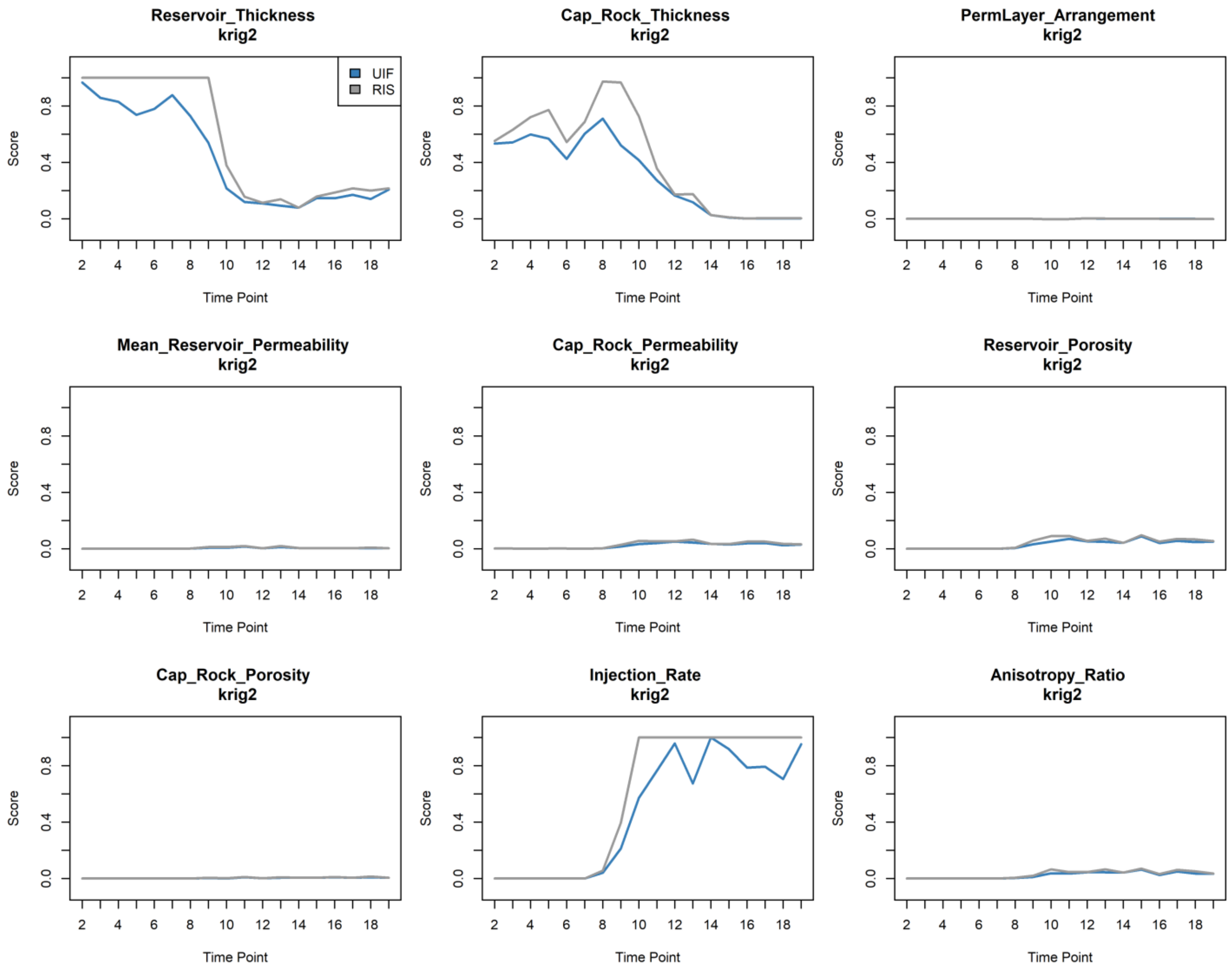


Figure 9-71. UIF vs. RIS for Remove.

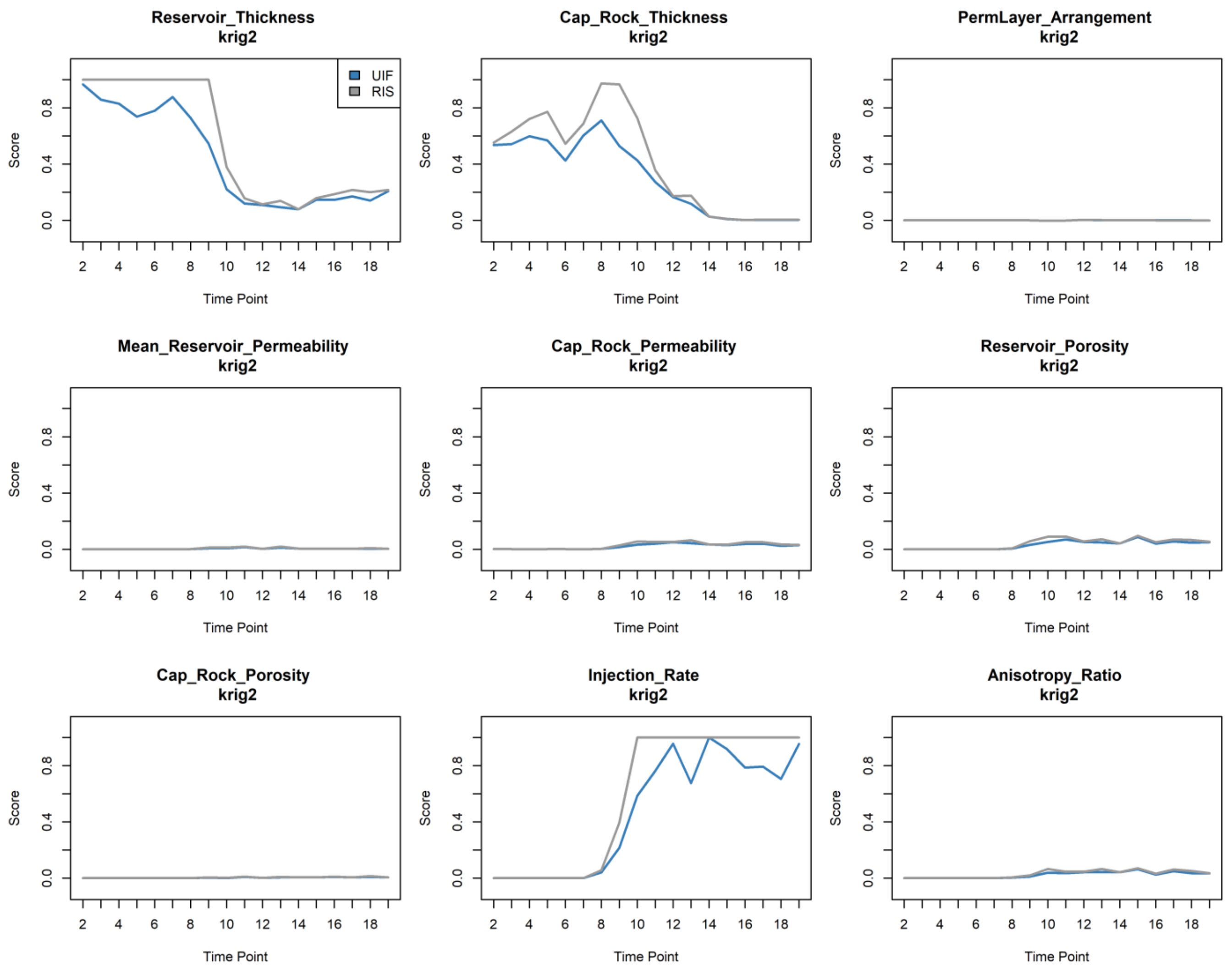


Figure 9-72. UIF vs. RIS for Permute.

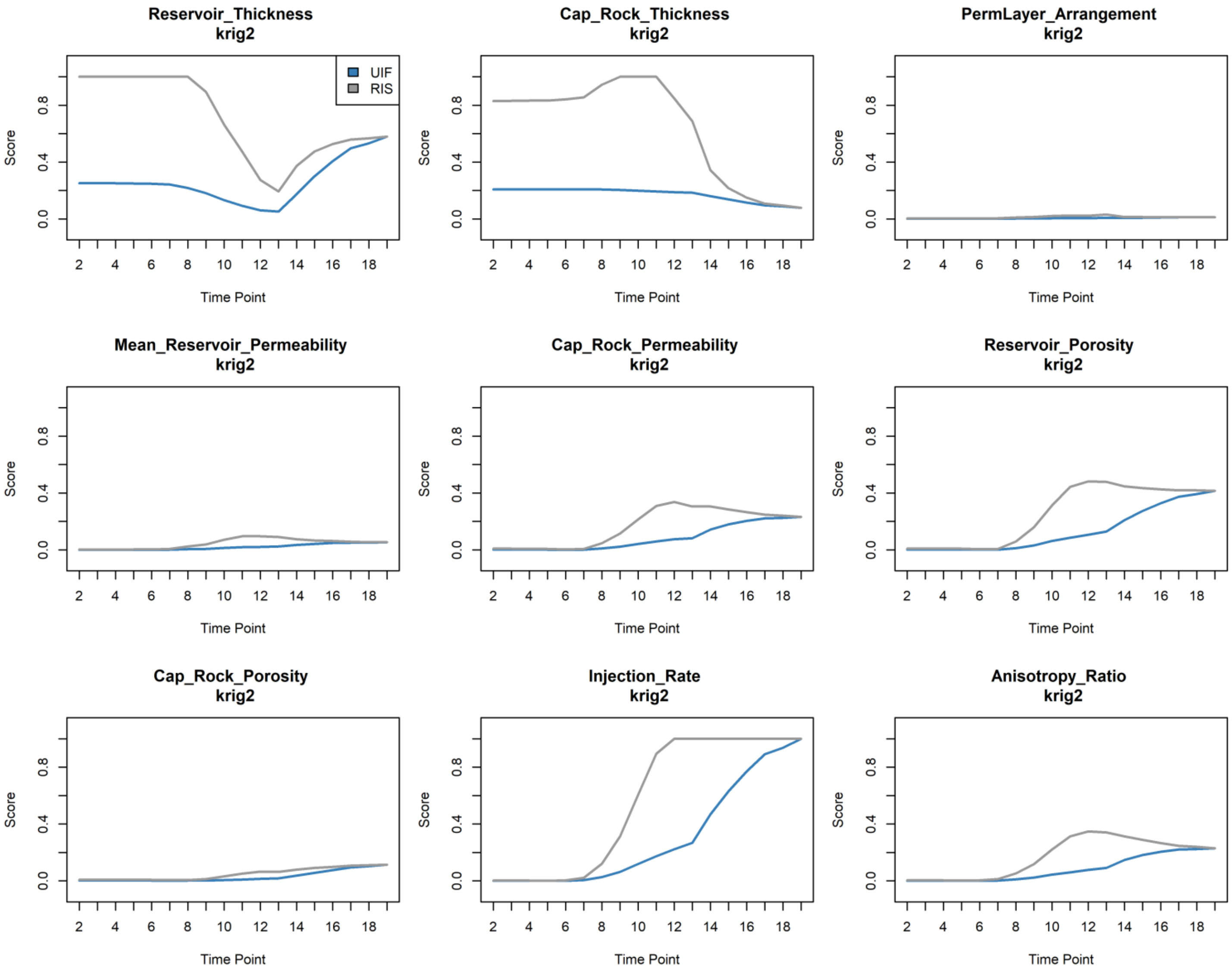


Figure 9-73. UIF vs. RIS for PDP.

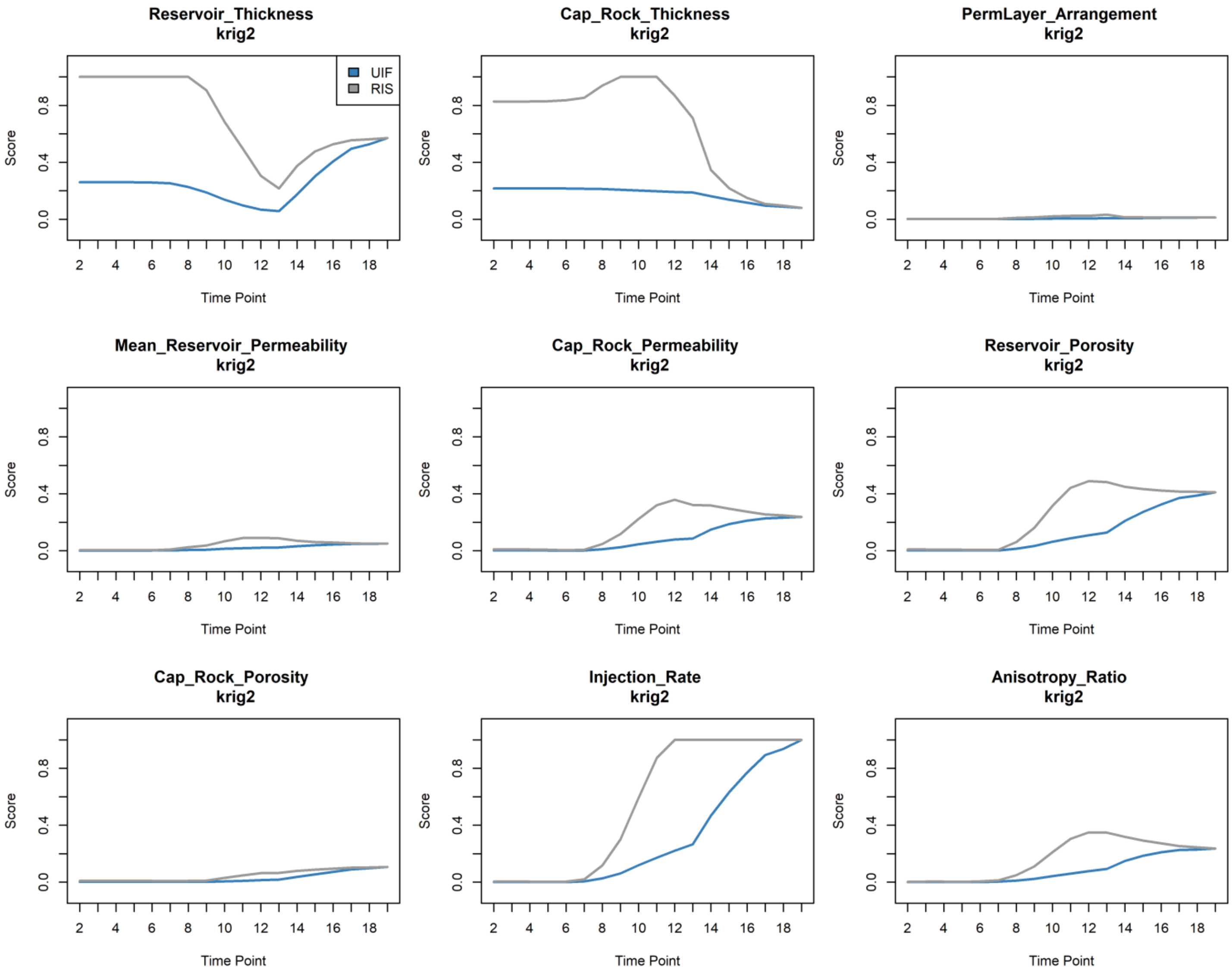


Figure 9-74. UIF vs. RIS for ALE.

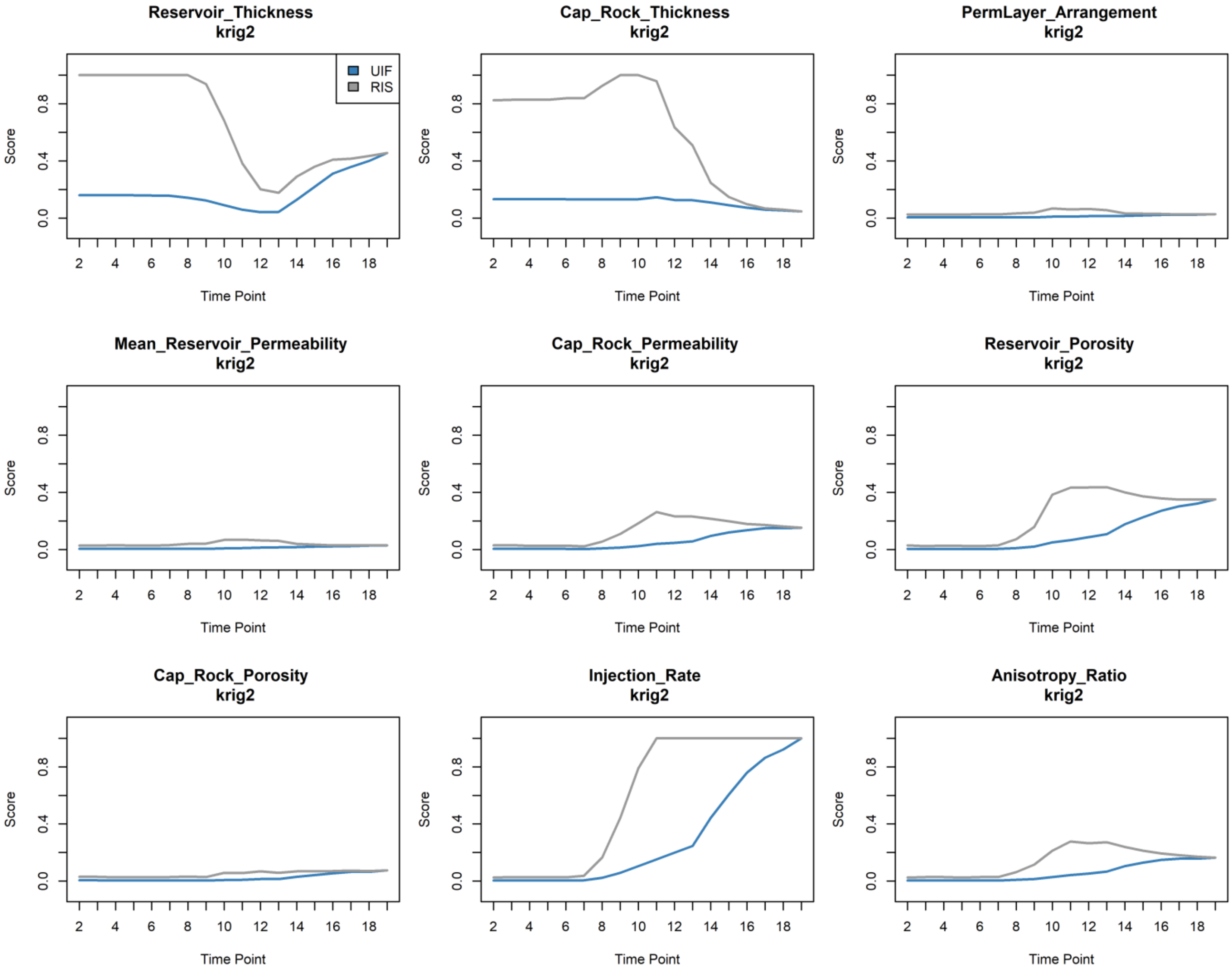


Figure 9-75. UIF vs. RIS for LIME.

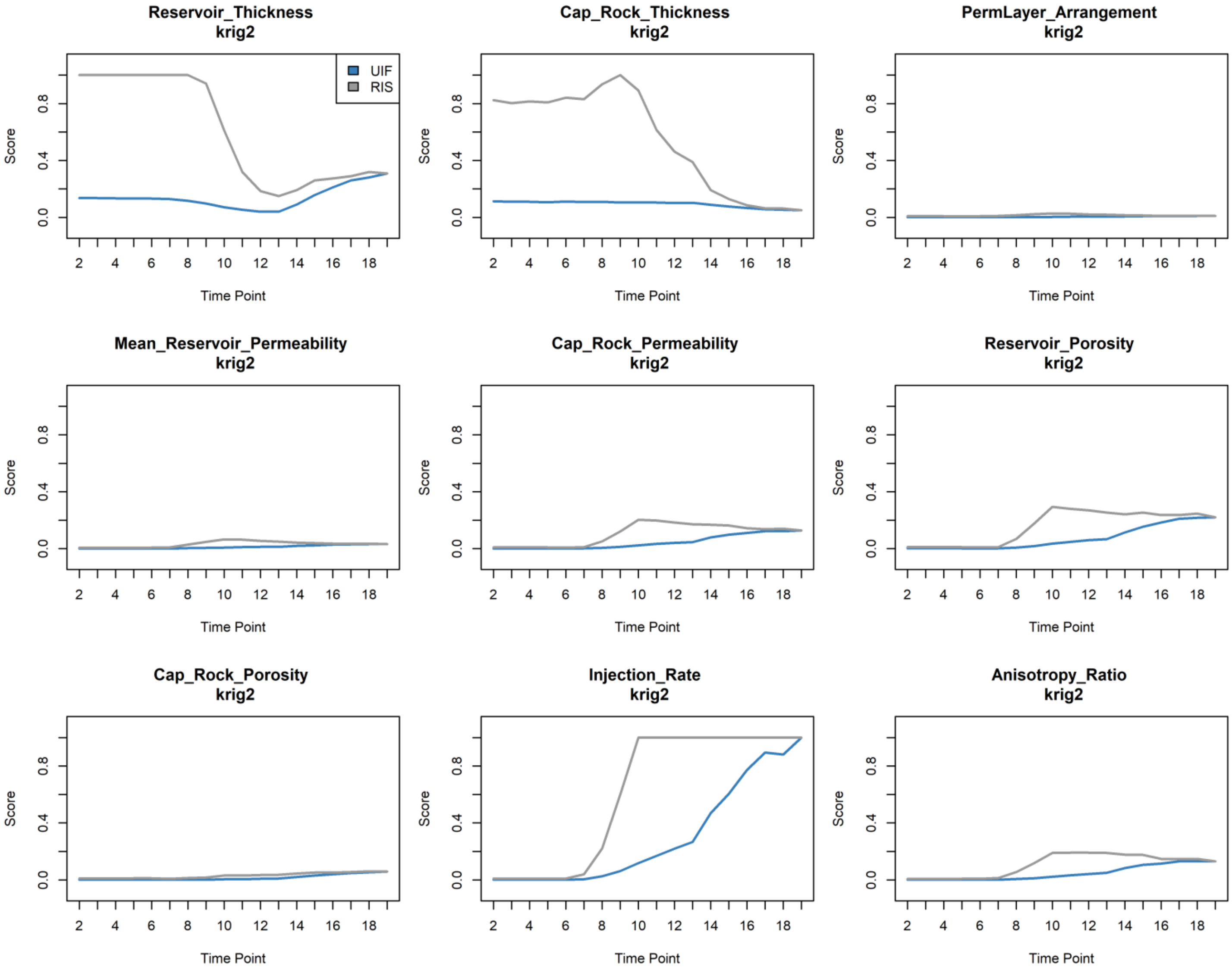


Figure 9-76. UIF vs. RIS for SHAP.

BATTELLE

It can be done