Merits of Pressure and Geochemical Data as Indicators of CO₂/Brine Leakage into a
Heterogeneous, Sedimentary Aquifer

Whitney Trainor-Guitton¹, Kayyum Mansoor², Yunwei Sun², and Susan Carroll²

¹*Colorado School of Mines, formerly of Lawrence Livermore National Laboratory*

²*Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, United States*

Submitted to

International J. Greenhouse Gas Control

NRAP Special Issue

ABSTRACT

This study assesses the merits of pressure data and geochemical data as indicators of a combined CO₂/brine leakage into a heterogeneous, sedimentary aquifer. We simulate the changes in three aquifer responses (pressure, total dissolved solids (TDS), and pH) due to CO₂/brine leakage at an abandoned well with an uncertain location and hypothesize that these changes can only be observed from a single shallow monitoring well, mimicking the low density of observation wells for the considered aquifer. Specifically, detection likelihoods are calculated to describe how frequently pressure, TDS, and pH signals will coincide with a leak for observations made at different distances and times from the initiation of the CO₂/brine leakage rate. The pressure signal gives a more spatially extensive signal than either TDS or pH, and pressure detection probabilities increase upstream of flow barriers (pressurizing-affect). The pH and TDS rebound down-stream of the flow barriers. When only considering the samples that experience the highest leakage volumes, there is a 50% likelihood of detecting a pressure change 400 m away at times ≥ 30 years. However, the TDS and pH detection likelihoods are $< 20\%$ at 100 m distance for times ≥ 30 years.

Keywords: CO₂ storage, CO₂ and brine leakage, monitoring, reactive transport, detectability.

1. INTRODUCTION

A critical component of carbon sequestration planning is monitoring of the area of review (or estimated impact area) to ensure non-endangerment of drinking water due to CO₂ and/or brine breaches from the underlying storage of CO₂. This has become known as Post-Injection Site Care or PISC (United States Environmental Protection Agency, 2012).

Many different monitoring techniques have been evaluated both synthetically and from the field to better understand the spatial and temporal movement of the CO₂ and pressure plumes within and above the reservoir. Gasperikova and Hoversten (2008) simulated the gravity anomaly created by the change of density due to the injection of CO₂ into a deep brine aquifer and a coal formation. Leuning et al. (2008) present a method for utilizing flux chambers to analyze atmospheric samples to detect if leakage has occurred. Others have interpreted physical changes due to CO₂ injection using measurements of electrical resistivity (Ramirez, 2005; Schmidt-Hattenberger et al., 2011; Dafflon et al., 2013; Trainor-Guitton et al., 2013; Yang et al., 2014, Yang et al., 2015), ground penetrating radar (Lassen et al., 2015) and seismic wave velocities (Mito and Xue, 2011). Daley et al. (2008) demonstrated how vertical seismic profiles give high-resolution images of the CO₂ distribution. Pressure and fluid saturations changes due to CO₂ injection has been monitored with 4D seismic (White and Johnson, 2009). Vasco et al. (2010) used interferometric synthetic aperture radar inversions to map a subsurface pressure plume associated with CO₂ injection. Other natural analogs or injection demonstrations targeting the shallow groundwater impacts used a suite of geochemical signatures (Evans et al., 2012; Frederico et al., 2004; Keating et al., 2009; Keating et al., 2014; Cahill et al., 2012; Peter et al., 2012; Spangler et al., 2010; Trautz et al., 2014).

Recent studies have quantified the accuracy of above zone pressure data to correctly identify a leak given that there will be natural variations in pressure within the shallow aquifer (Sun et al., 2013). They found that pressure detection was related to the correlation length of aquifer flow properties as well as the distance between the leakage source and the monitoring well. However, a large uncertainty will be the location of the leaking well and in turn its distance to the monitoring well. The distance between the leak source and the nearest monitoring well may deem above zone pressure an unviable detection method, especially if the monitoring well

density is quite sparse. This is true for two significant regional aquifers in the United States: the Ogallala (Kansas, Nebraska, Texas) and Edwards (Texas) aquifers with 1 to 2.6 wells per km² (Carroll et al., 2014). It is important to determine the efficacy of the different possible monitoring techniques with uncertain leak locations to design optimal monitoring protocols to ensure non-endangerment to drinking waters. Keating et al. (2014) analysed and compared synthetic groundwater sampling and pressure monitoring for the unconfined, carbonate Edwards aquifer, concluding that the probability of detection is quite low due to the unconfined conditions, the low monitoring well density compared to the plume size and assumption that the leak location is unknown.

The purpose of this study is also to evaluate and compare the reliability of above-zone pressure and groundwater solution chemistry monitoring as leak diagnostics. We generate synthetic pressure, total dissolved solids (TDS) and pH data via many subsurface realizations and leak simulations that are based on a specific part of the confined High Plains Aquifer, also known as Ogallala, with an average well density of 1 per km² (Carroll et al., 2014). We use a simulation approach to explore how the following principal uncertainties influence the three signals to correctly diagnose a CO₂/brine leak:

- **The distance between the source leak and the monitoring well.** Although it is likely that instrumentation will be around the injection well at many depths, it is still possible that unknown abandoned wells or faults could serve as random leak locations thus unknown *a priori*, as demonstrated in Gal et al. (2013). This study explores the likelihood of detection given that the distance between the monitoring well and leakage source could vary between 50 and 900m.
- **The heterogeneity of the aquifer flow properties.** This study uses two-facies heterogeneity models of High Plains Aquifer (Carroll et al., 2014) to evaluate the effects of heterogeneity on spatial and temporal scales of the signals.
- **The leakage rates of CO₂ and brine.** The magnitude and duration of the leak will be one of the main determiners of the extent of the three signals. **We utilize a characteristic flux model to capture a range of possible CO₂ and brine leakage fluxes into the shallow aquifer, due to varying reservoir and wellbore conditions.**

We aim to use these results to estimate the efficacy of the three signal types, and ultimately define a framework that can inform operators and regulators on the appropriate spatial and time scales for operational and post-injection site care (PISC). We base our analysis on the detection

1 likelihoods that are computed from our simulation. The detection likelihoods define how
2 frequently pressure, TDS, and pH signals will coincide with a leak. We do not consider how
3 often false positives could occur; this would require either 1) explicit simulation of pressure
4 disturbances from pumping, natural recharge or other activities near the monitoring well or 2)
5 incorporating statistics of how large and how often this kind of “noisy” pressure signal would
6 occur. Wang and Small (2014) include false positives in their single-phase, analytical study of
7 CO₂ leakage via seepage through the caprock by assuming measurement error is characterized by
8 a lognormal distribution.

9 Through this analysis we focus on:

- 10 • Recommendations for early leak detection. It is important to detect leaks to drinking
11 water as early as possible because small amounts of leaked CO₂ and brine can change
12 groundwater chemistry for long periods of time (Carroll et al., 2014).
- 13 • Confidence that the absence of signal represents the absence of a leak. US EPA
14 guidelines are designed to be flexible; such that duration of post injection site care (PISC)
15 can be reduced if the operators can show that Underground Sources of Drinking Water
16 (USDW) are not endangered from the stored CO₂. This may be difficult when impacted
17 volumes are small, however, it may be possible to infer the upper volume threshold in the
18 absence of signal over a specific monitoring window.

19 The following Methodology section will describe the geologic models representing the
20 aquifer heterogeneity, the simulations that generate the pressure and geochemical signals and
21 finally the calculations of detection likelihoods. The next section will present the results, and
22 the last section will discuss these results.

2. METHODOLOGY

2.1 THE MODEL ARCHITECTURE & THE UNCERTAIN PARAMETERS SAMPLED

The High Plains Aquifer in Southwest Kansas represents an important and vast groundwater resource. We consider how a leaking well could introduce CO₂ and brine from a deep storage reservoir and affect the water quality of an aquifer volume of 1 km × 1 km × 240 m within the High Plains Aquifer. Our models represent this volume with 40 × 40 × 48 cells, such that each cell represents 25 m × 25 m × 5 m in the x, y and z directions. The location of the leaking well is fixed at $x = 600$ m, $y = 350$ m and $z = 0$ m (bottom of aquifer). This gridding and placement of the leaking well allow for observational distances of 5 m to 990 m from the leaking well to the monitoring well. Fig. 1 displays the distribution of all the possible observation distances given the fix location of the leak. Given the well density of the High Plains Aquifer, it is a reasonable assumption that only 1 monitoring well would exist within this volume.

Fig. 1. Distribution of observation distances for potential monitoring wells with pressure transducer and geochemical sampling, given the gridding of the simulation

We are particularly interested in how the heterogeneity of the High Plains will affect the detection likelihoods of pressure and geochemical signals. We varied the correlation lengths of the vertical and horizontal permeability for the aquifer and the sand proportion from a uniform distribution with the minimums and maximums given in Table 1. Lithologic information within this area (from irrigation, public water supply and oil field water supply wells) were used to

develop the geostatistically derived indicator models using the TPROGS geostatistical algorithm (Carle, 1999; Carroll et al., 2014). Through the parameter ranges of Table 1 and the stochastic TPROGS algorithm which utilizes transition probabilities between facies, we have 500 realizations of the heterogeneity: discrete, geologic bodies that represent sand and clay facies (see Figure 2). These two lithologies are presumed to have permeability's that are orders of magnitude different: 10^{-11} m^2 for sand versus 10^{-18} m^2 for clay, which were taken from the USDA Rosetta database (Schaap et al., 2001). This heterogeneity may be more representative of actual subsurface conditions than previous studies where the hydraulic conductivity was a continuous variable (Sun et al., 2013). Note that the maximum horizontal length of flow barriers (2500 m, Table 1) can be larger than the uniformly spaced part of the simulation grid (1 km). We will discuss later how this may influence our results.

Table 1: Minimum and maximum values for aquifer heterogeneity parameters

Parameter	Minimum	Maximum
Horizontal sand/clay correlation length	200 m	2500 m
Vertical sand/clay correlation length	5 m	25 m
Sand volume fraction	0.35	0.65

Figure 2: Example heterogeneity of aquifer produced by TPROGS; red indicates sand layer, blue represents clay layer

This aquifer study represents the central part of previous work that represented a 5 km by 10 km by 240 m volume of this aquifer to evaluate the long-term (200 years) risks of geochemical alterations due to possible CO₂ and brine leaks (Carroll et al., 2014). Unlike the regional model of Carroll et al. (2014), no regional flow is imposed on this smaller, higher resolution subsection.

Fig. 3. High Plains Aquifer model location and dimensions. The simulation grid used for this study subsamples spatially and temporally from the original study (Carroll et al., 2014).

2.1.1 Boundary Conditions of Leak Simulations

Closed boundary conditions are imposed on the bottom and sides of all of the simulations. To avoid boundary condition affects, the model domain encompasses a $3 \text{ km} \times 3 \text{ km}$ area, where the outer 2 km in both the easting and northing direction are considered grid padding. Specifically, the grid padding consists of five cells of increasing size: 100, 150, 250, 500, and 1000 meters. The heterogeneity continues into these cells, and the pressure and geochemical response can extend into these cells, however, they are not considered in the detection likelihood calculations. Gas is allowed to escape to the atmosphere (the top of the model). The bottom boundary condition of the model is closed except for the leak. The model is fully saturated up to the upper 50m where it is partially saturated. Since no fluxes are removing water (pumping water) this remains constant over the 5-year simulation.

2.1.1.1 Characteristic Leakage Flux Model

Previous studies have demonstrated the dependence of leakage fluxes on CO_2 injection rate, injection duration, reservoir permeability, distance between the injector and leaky well, etc. Due to the uncertainties of the geologic formation and system parameters, it is computationally expensive to simulate all possible scenarios with high-fidelity models. Therefore, we utilize a characteristic flux model, which borrows CO_2 and brine saturations and pressures from separate physical models of the reservoir and wellbore to mimic the patterns of the CO_2 and brine that would enter the shallow aquifer. We describe in detail how we achieve this.

First, we use the CO₂ saturation and pressure distributions modeled by Wainwright et al. (2013) which simulated CO₂ injection into 200 realizations based on the Kimberlina reservoir site. These realizations captured the uncertain reservoir permeability and porosity, and caprock permeability. Next, following Lu et al., (2012), the Kimberlina pressures and High Plains hydrostatic pressures are used as input into 48 different wellbore models that conceptually represent possible abandoned wellbores that may provide communication between the Kimberlina storage reservoir and the overlying High Plains aquifer. For each of the 48 wellbore models, 1,000 realizations are derived, which consider the uncertainty in well depth, well permeability and the CO₂ saturation and pressure derived from the Kimberlina results. The well permeability spans between 1.0e-12 and 1.0e-10 m². The outputs of the wellbore models are CO₂ and brine fluxes into the High Plains Aquifer.

These CO₂ and brine leakage fluxes, from physical reservoir and wellbore models, can be mimicked using trapezoid and terraced shape functions of time, as illustrated in Figure 4, and characterized by geometric parameters. Therefore, the characteristic flux model becomes more general for a range of CO₂ injection periods and wellbore permeability. The characteristic flux model is a surrogate but not a reduced-order model. Using the patterns and physical-model results, we calculated uncertainties (ranges) of characteristic parameters that are used as uncertain inputs of our flow and transport models (rather than polynomial flux functions of uncertain parameters as in the case of reduced-order models). Therefore, the maximum CO₂ flux reflects the leak occurrences when the CO₂ injection rate, reservoir permeability, and the distance between the injection and leaky well culminate in the highest CO₂ flux. Similarly, the brine and CO₂ flux of our characteristic flux model (Fig. 5) are correlated since they are derived from physical models of multiphase flow.

Conceptually, CO₂ flux increases to a plateau during the onset of CO₂ injection into the storage reservoir, maintains a constant flux during injection, and then decrease overtime to some minimal level. Variability in the CO₂ leakage profile was described using four parameters:

- q_{CO_2} – the peak flux,
- T_{1C} – the time needed to reach peak flux,
- dT_{2C} – the duration of the peak flux, and
- dT_{3C} – the duration of the transition to zero flux after injection has stopped.

Brine leakage profiles are different from that of CO₂. Brine leakage was characterized with a maximum and constant flux during injection, which falls off to a final, non-zero flux after injection stops. Uncertainty in the brine leakage profile was generated using four parameters:

- q_{BRN} – the initial and maximum flux
- λq_{BRN} – the final flux
 - λ : Ratio = minimum and maximum of brine flux; brine leakage never reaches 0.
- $T_{1\text{B}}$ – the injection time, and
- $dT_{2\text{B}}$ – the duration of the transition between the maximum and final flux.

Figure 4: Schematic of the CO₂ and brine leakage model parameters and profiles in the generalized model.

Table 2: Parameter ranges for generalized CO₂ and brine leakage models

Parameter	Min	Max	Units
q_{CO_2}	0.001	0.5	kg / s
q_{BRN}	0.005	0.075	kg / s
λ	0.2	0.3	Ratio for brine-leakage tail
$T_{1\text{c}}$	5	50	yr
$dT_{2\text{c}}$	0	100	yr
$dT_{3\text{c}}$	5	50	yr
$T_{1\text{b}}$	1	50	yr
$dT_{2\text{b}}$	1	10	yr

Using the parameter ranges of Table 2, Fig. 5 displays the *cumulative* leakage flux magnitudes for the 50 year simulations for 374 successful simulations.

Fig. 5. Cumulative leakage of CO₂ (top) and brine (bottom).

2.2 GENERATION OF SIGNALS & DETERMINATION OF LEAK

The objective of this study is to evaluate how reliable above zone pressure monitoring and groundwater sampling would be as leak diagnostics. We generate these synthetic signals by sampling the parameters described above to represent CO₂ and brine leaks, and then simulating how the CO₂ and brine will generate pressure and geochemical changes in the shallow aquifer. This section describes how we establish what is a detectable signal and what classifies as a leak. These are ultimately used to determine how frequently pressure, TDS, and pH signals will coincide with a leak. We will refer to this as the detection likelihood.

The pressure and groundwater chemistry signals for all samples are generated by concurrently varying the following parameters:

- horizontal and vertical correlation length for:
 - sand facies
 - clay facies
- sand volume fraction
- time-varying flux magnitudes of:
 - brine
 - CO₂

First, 500 subsurface models are generated with the ranges described in Table 1; an initialization (steady-state) was run for 1,000 years for each sample to generate the initial conditions specific to each sample, where only native, fresh (background) water is in the aquifer with no fluxes imposed. Carroll et al., 2014 describe the background conditions of the aquifer. Next, each sample is paired with a unique CO₂ and brine leaking flux profile, which are generated from the leakage scenarios shown in Fig. 5. All samples simulate 50 years; 374 of the 500 samples completed successfully. The samples that did not complete after 72 CPU hours had flow barriers in proximity to the leakage flux that cause convergence issues.

2.2.1 Pressure Signal

Each leak scenario is simulated using LLNL's NUFT simulator (Nitao, 1998). Each sample's CO₂ and brine leakage rate is imposed as a flux on the bottom of the aquifer to simulate a leaking wellbore. The CO₂ flux is purely in the gas-phase (since the leakage occurs at a shallow depth of 240m and temperature of 15°C), and the brine flux is in the liquid phase. NUFT determines the pressure for each grid cell for all time steps. Fig. 6 shows an example of the *relative* (not absolute) pressure change between the time step of 55 days and 60 days. The relative pressure change allows for better visualization of the patterns of how pressure is moving through the aquifer. However, for signal generation, δp (change in pressure) will always be absolute (compared to time=0: t_0).

Fig. 6. Relative pressure change (60 days versus 55 days); color bar represent Pascals.

The accuracy and precision of pressure transducers that will measure pressure changes in-situ depend on the field detection limit of a specific transducer. We define a detection limit of 0.3 PSI (0.002 MPa), which is appropriate given the depths of the modeled aquifer and within ranges used in previous studies (In-Situ, 2014; Nogues et al., 2011; Keating et al., 2014). This study does not consider situations where the increase of 0.3 PSI or greater maybe not be related to a leakage event (i.e. a false positive). False positives could be included in a future study if statistics about pressure increases unrelated to leaks were available, such as from field observations of pressure data. Sources of noise that would create the false positives would need to be added to the simulations. The pressure signal, denoted by δp , is computed by calculating the absolute change in pressure from time = 0 (t_0) to time = t for each cell (k) within the grid for every simulation, s

$$\delta p_{sk}(t) = P_{sk}(t) - P_{sk}(t_0) \quad \forall s, k, t. \quad (1)$$

2.2.2 Groundwater Chemistry Changes

Each simulation includes reactive transport capabilities to account for changes in groundwater pH due to CO₂ dissolution, as well as dissolved sodium, chloride, calcium, and bicarbonate to calculate the TDS. NUFT simulates the phase transitions of CO₂. The dissolution of CO₂ in groundwater stimulates the following sets of reactions:

These reactions promote the acidification of the system, which is then buffered by calcite dissolution. Trace metal concentrations, which can be introduced to the system due to pH-related changes or originating from the brine, are not considered in this study.

The changes in groundwater chemistry are used to determine the TDS and pH signals and to define the presence of a leak: the volume of groundwater that exceeds certain water quality thresholds. Like Carroll et al. (2014), we considered a “no-impact” threshold for the High Plains aquifer. Developed by Last et al. (2015), the no-impact threshold represents the lowest detectable concentrations above the background water chemistry that could be used to quantify a change in groundwater chemistry due to CO₂ or brine leakage. For the High Plains aquifer, this was determined to be 1,300 mgL⁻¹ and a pH of 7.1. We use a slightly more acid pH of 6.625, because higher pH did not yield reliable reduced order models from the numerical data sets (Carroll et al., 2014). Specifically, a simulation at time t is defined as a leak ($\theta_s(t) = 1$) if *any voxel* ($\exists k$) representing 2,000 m³ of aquifer exceeds 1,300 mgL⁻¹ TDS or a pH 6.625:

$$\theta_s(t) = \begin{cases} 1, leak & \text{if } \begin{matrix} TDS_{sk}(t) \geq 1300 \\ pH_{sk}(t) \leq 6.625 \end{matrix} \\ 0, no leak & \text{otherwise} \end{cases} \quad \exists k, \forall s. \quad (5)$$

1 Fig. 7 plots the percentage of samples diagnosed as a leak per time step. Roughly 60% of the
 2 samples are deemed a leak by the second year.

3

4 **Fig. 7. Percent (%) of samples diagnosed as leak per time step, using $1,300 \text{ mgL}^{-1}$ as threshold of TDS (for any**
 5 **voxel).**

6 **2.3 CALCULATION OF LIKELIHOOD OF DETECTION**

7 This section describes how we calculate the detection likelihoods, which describe how frequently
 8 pressure, TDS, and pH signals correctly identify a leak. We will describe this with the pressure
 9 variable (δp), but the same approach is used with the TDS and pH data. A Boolean count
 10 ($c_{dp_{sk}}(t)$) is made to assess which locations within the model and at what time they exceed the
 11 given threshold, where the thresholds relate to the accuracy of pressure transducers for shallow
 12 groundwater aquifers or the background chemistry of the groundwater. This is done for each cell
 13 (k) within the grid for every simulation, s , and at every time step:

$$c_{dp_{sk}}(t) = \begin{cases} 1 & \text{if } \delta p_{sk}(t) \geq \text{threshold} \\ 0 & \text{if } \delta p_{sk}(t) < \text{threshold} \end{cases} \forall s, k, t \quad (6)$$

1 Next, the signals at different locations within the model domain (k) are grouped by distances
 2 from the leak location (d). The distances range from 5 m to 990 m and are binned in intervals of
 3 5 m. By grouping the signals into distances (for all samples), we can infer the likelihood of
 4 observing a signal ($c_{dp_s=1}$) at certain distances from a leak with an unknown location.

$$c_{dp_s=1}(d, t) = \sum_{k=1}^K c_{dp_{sk}}(t) \quad \text{where } \|k - \text{leak}\| \in d \quad (7)$$

5 The counts of Eq. 7 can be screened by their absolute location in the model domain (represented
 6 by k). We calculate these counts theorizing the location of the observation well could be located
 7 1) anywhere within the model domain, 2) above or 3) adjacent the leak location. These three
 8 possible observation volumes are represented in 2D in Fig. 8. The yellow, orange and red boxes
 9 use all locations (k) within the model, only those vertically above the leak and only those radially
 10 separated from the leak to count signals and non-signals, respectively.

Fig. 8. Example vertical cross-section of permeability; detection likelihoods are computed assuming an observation well lies within one of the three squares: yellow, orange (only vertical from leak) or red (only horizontal).

With this evaluation (Eq. 7) and the determination of a leak (Eq. 5), the joint occurrence of both being a leak and signal occurring at time t is computed

$$c_{dp=1,leak=1}(d,t) = \sum_{s=1}^S c_{dp_s=1}(d,t) \text{ if } \theta_s(t) = 1 \quad (8)$$

and the occurrence of both no signal and a leak occurring:

$$c_{dp=0,leak=1}(d,t) = \sum_{s=1}^S c_{dp_s=0}(d,t) \text{ if } \theta_s(t) = 1 \quad (9)$$

1 Thus, the detection likelihood for particular distances (d) and times (t) can be calculated by
 2 normalizing each count by the total number of occurrence of either signal ($c_{dp=1}$) or no signal
 3 ($c_{dp=0}$):

$$Pr(\Delta P = \delta p(d, t)_{signal} | \Theta = \theta(d, t)_{Leak}) = \frac{c_{dp=1, leak=1}(d, t)}{c_{dp=1, leak=1}(t) + c_{dp=0, leak=1}(d, t)} \quad (10)$$

4 The normalization in Eqs. 9 and 10 will depend on which locations within the model have been
 5 determined as possible monitoring well locations (represented by colored squares in Fig. 8).
 6 Finally, the detection likelihood (Eq. 10) assumes that the length of a screened observation well
 7 will be at least 5 m (~15 feet) because of the dimensions of the grid cells and the distance
 8 intervals used.

3. RESULTS & OBSERVATIONS

This section presents the resulting likelihoods of the pressure, TDS and pH signals to effectively indicate CO₂/brine leakage. The baseline thresholds of 0.3 PSI (0.002 MPa), 1,300 mg/L and $10^{-6.625}$ H⁺ are used as the minimum measurable change of pressure, TDS and pH, respectively. Fig. 9 presents the statistics of the cumulative impact (how large the geochemical leak signatures are) by time in terms of TDS (the pH leak signatures are comparable). The box plots of Fig. 9 display the mean (red line), the interquartile range, or IQR, (solid blue box), 1.5 times the IQR (dashed blue) and extreme values (red circles).

Fig. 9. Statistics of cumulative impact for all samples diagnosed as a leak ($\theta(t)=1$).

3.1 DETECTION LIKELIHOODS BY DISTANCE AND TIME: ACCORDING TO ALL SAMPLE SIMULATIONS

The detection likelihoods (probabilities) by distance (relative to source) and time are calculated first assuming that the observation well could be at any location with the 1 km × 1 km × 240 m domain. This domain is enclosed by the yellow square in Fig. 8 (representing an example vertical

cross section), which also indicates the leakage location. The detection likelihoods for the yellow square are shown in Fig. 10. The top plots contain the detection likelihoods for 8 time slices that are less than 2 years (730 days) while the bottom plot contains the likelihoods for 8 time slices between 2 through 48 years. Most notable is that the TDS and pH signals do not extend beyond 300m distance from the leakage source. In contrast, the pressure signal is present in 20% of the samples at 840m from the leakage source (farthest possible from leakage point in model domain) 30 years after the leakage flux is introduced.

Fig. 10. Considering signals within entire model domain: detection likelihoods by distance & time for TDS
1300 leak definition and baseline signal thresholds.

The “kinks” in the likelihood of Fig. 10 can be understood if the detection likelihoods are calculated for vertical and horizontal locations relative to the leakage source. First we consider only the observation locations that are directly above the leak (i.e., all cells with same x and y as leakage location). This is represented by the orange square in Fig. 8. Fig. 11 contains the detection likelihoods for these locations. The pressure signal in Fig. 11a shows which oscillations coincide with the maximum vertical correlation length of the facies. The likelihood of observing a pressure disturbance increases as the vertical distance approaches 25 m, reaching a high of 70% for times ≥ 12 years. The likelihood then dips to 30% at 50 m, followed by

another increase at approximately 75 and 115 meters above the leak. These oscillations represent how the pressure signal physically builds behind the flow barriers, and then dissipates after alternate flow paths around these barriers are found. This is seen in Fig. 12c) for sample 534 where the contours represent the locations that exceed the detection threshold of pressure change at particular times. There are two distinct pressure bodies with centers at 25 m and 60 m above the leak, which are located under the two deepest flow barriers (Fig. 12d).

Fig. 11. Considering only signals directly vertical of leaking well.

In contrast, the TDS signal appears to rebound after the maximum vertical correlation length. At times ≥ 12 years, the signal rebounds from 10% to 20% likelihood of detection from 25 m to 30 m above the leak. This can be explained by the different time scales of the TDS versus pressure signals. The alternate flow paths that move beyond the flow barriers (>25 m) are already established by the time the ion-rich water reaches these barriers. The velocity of the flow is likely faster before and slower after the flow boundary, thus allowing for both accumulation of TDS and geochemical reactions that contribute to the TDS signal (Eqs. 2-4). This is visually seen in Fig. 12b where the TDS contours have vertical conduit features below and between the dark flow barriers (depicted in Fig. 12d).

The pH signals rebound similarly as the TDS signal after the vertical boundary, but at earlier times: 4 years versus 12. This is reasonable since the pH signal is largely driven by the movement of CO₂ through the subsurface, which travels more quickly than the brine due to buoyancy. Additionally, this rebound is experienced closer and closer to the leak at later times. For example, at 30 years, the increase in detection likelihood is at 20 m (instead of 25m) above the leak. This could indicate how the geochemical reactions occur in other directions other than the principle flow direction. Both the TDS and pH also demonstrate a noticeable rebound after 50 and 100 vertical meters from the leak, at the second and third correlation length of heterogeneity.

Fig. 12. Sample 534 contours of a) <6.625 pH b) >1300 TDS c) >Δ0.3 PSI (0.002 MPa) and d) log₁₀ permeability.

1 The detection likelihoods are calculated for observations that are only radially offset from the
2 leak to understand further how heterogeneity affects the different signals. The red square in Fig.
3 8 illustrates the area of these observations; the likelihoods produced by only considering
4 observations in this area can be used to theorize how likely a signal would be detected in a
5 monitoring well that is within the same geologic unit (same approximate depth) but located at
6 some distance radially from the leak. Fig. 13a demonstrates how the pressure signal generally
7 propagates uniformly across the domain: the likelihood of detection is equivalent for distances
8 between 20 m and 500 m. By year 2, the pressure signal is present in almost 20% of the samples
9 that are classified as a leak. This likelihood peaks at 42% at 30 years, dropping to 38% at year
10 48. The TDS and pH signal, however, only extend beyond 100 m in 10% of the samples. The
11 TDS signal is only present in 20% of the samples at 50 m and ~5% at 100 m, however, it is
12 highest at 48 years for all distances, indicating that perhaps the extent of TDS signal could
13 increase with longer observation times. The pH signal is also seen in about 50% of the samples at
14 50 m radial distance, and also like TDS, pH signal is always most likely at year 48. Unlike TDS,
15 the pH signal reaches at 50% earlier: 30 years. Additionally, the pH signal is seen in just under
16 10% of the samples at 200m away from the leak, reaching radially further than the TDS signal.

Fig. 13. Considering only signals radial from leaking well (within 10m of the same depth of the leak).

3.2 DETECTION LIKELIHOODS: SIMULATIONS WITH TOP 25% LEAKAGE RATES

As described previously, the definition of leakage in all samples is related to the protection of USDW standards: if any volume of aquifer has been altered then a leak has occurred. Thus, if one voxel reaches or exceeds 1300 mg/L TDS or pH is <6.625 , the sample is considered a leak and is used in calculating the detection likelihoods (Fig. 10-Fig. 13). If the “true” extent of the leak is only 2,000 m³ (the volume that one voxel represents) then it is reasonable that the detection likelihoods would be low at many observation distances as is illustrated in Section 3.1. We examine the actual size of the impacted groundwater for the high CO₂ and brine leakage cases to reveal how effective the pressure, pH and TDS signals are at indicating larger leaks.

We consider only the samples that experience the top 25% CO₂ and brine cumulative leakage flux by year 10. As illustrated in Fig. 5, the CO₂ and brine cumulative fluxes are of comparable magnitude for all samples until year 10; then, the CO₂ cumulative flux surpasses the brine flux one-fold every 10 years, such that by 50 years 5 times more CO₂ has entered each sample than brine. Fig. 14 displays the statistics on the volume of aquifer exceeding 1300 TDS for the samples experiencing the top quartile of cumulative CO₂ leakage. The box plots display the mean (red line), the IQR, (solid blue box), 1.5 times the IQR (dashed blue), and extreme values

1 (red circles) for the same time intervals of the previously shown detection likelihoods (Fig. 10-
2 Fig. 13).

Fig. 14. Statistics of cumulative impact for samples with top quartile CO₂ leak.

Fig. 14 demonstrates how the cumulative impact (in terms of TDS) greatly exceeds the leakage definition of 1 voxel (2,000 m³) at 2 years for the higher leakage samples. Additionally, it is observed that the rate of increase in the affected aquifer slows after 20 years, which is consistent with the leakage scenario imposed on the samples (Figure 4). At 48 years, the mean aquifer volume that exceeds 1,300 TDS is approximately 10⁷ m³, which represents about 5,000 voxels. However, detecting the leak with a single observation well that could be anywhere within the 1km by 1km by 240 m domain (2.4 × 10⁸ m³) will still be challenging because the mean leakage-impacted volume of 10⁷ m³ only represents ~4% of the model domain (2.4 × 10⁸ m³).

Fig. 15 contains the detection likelihoods for these samples that have the top quartile CO₂ cumulative rates (Fig. 14). Compared with the detection likelihoods of Fig. 10, the detection

likelihoods for the pressure signal are higher at all times for samples with the highest 25% of CO₂ cumulative flux. Twelve years after the leak, 60% of the samples categorized as leak have a pressure signal 100 m from the source, compared with 30% when including the smaller leaks. The pressure signal is present in 30% of the samples at distances >300 m of the leak for higher leak fluxes compared with only 10% presence for all simulations (Fig. 10).

Fig. 15. Detection Likelihoods calculated only with samples with top 25% CO₂ cummulative leakage (by 10 years).

When filtering for only higher CO₂ leakage, the TDS and pH signals are more likely to be detected. However, they are still spatially limited compared to the distances at which pressure changes are observed. The TDS signal likelihood increases to 10% for the higher leakage samples at 12 years, compared to ~2% of all the samples at 100m from the leak. The detection of a pH signal at 100m is ~15% versus ~5% when considering only high leakage samples. The likelihood of observing a TDS and pH signal drops further with distance with only a 5% likelihood at ≥ 300 m.

Fig.16. Statistics of cumulative impact for samples with top quartile brine leak.

Next, we consider the samples that experience the top quartile of brine leakage volume. Fig.16 contains the statistics of the cumulative impact volumes for these samples. Compared with Fig. 14, the mean volume of affected aquifer is close to an order of magnitude less at 48 years. This reflects how the brine leakage flux is simulated to decrease earlier than the CO₂ flux (Figure 4).

Fig. 17 contains the detection likelihoods considering only the samples with the top quartile of cumulative brine leakage rates (at 10 years). Considering only these samples, the likelihood of a pressure signal is markedly increased at 2 years for most distances. For later times (e.g., 21-50 years), the uptick at 75 m reaches ~55% likelihood of detection versus 45% likelihood when including all leakage rates. But unlike the high CO₂ volume (Fig. 15), for the samples with high brine volume (Fig. 17), the likelihoods for all three signals beyond 100 m look comparable to those using all samples (Fig. 10).

1 **Fig. 17. Detection Likelihoods calculated only with samples with top 25% brine cummulative leakage (by 10**
2 **years).**

4. CONSIDERATIONS FOR MONITORING USDW ABOVE CO₂ STORAGE SITES

From these detection likelihoods we can make certain recommendations for monitoring plans and post-injection site care (PISC). Firstly, effort and resources should be made in the thorough placement of pressure transducers, because pressure signals can give a more spatially extensive and consequently, a more timely indication of a leak if the monitoring well is >100 m away, especially when compared with point source chemical analysis. Despite the spatial range that pressure provides for some situations, many pressure disturbances in the groundwater due to leaks may be too small to be detected. The detection likelihoods suggest that the pressure signal has an even probability of detecting leaks that produce plumes around 1,000,000 m³ (or 0.1% of one cubic km³) when using 1 monitoring well per km² (the average for the High Plains aquifer, Carroll et al., 2014). The probability of detecting leaks would increase with increasing plume size.

Secondly, point source monitoring locations require a density greater than the 1 per km² used in this study to gain confidence that “non-detects” indicate the absence of impacted groundwater. We see 50% likelihood of detecting a pressure change 400m away after ≥30 years for the top quartile of CO₂ and brine leaks. The need for multiple monitor wells is especially true in early times where plumes from even the top leaks are small and have low likelihood of detection (less than 5 % for 4 years or less when monitoring wells are within 100 to 400 m of the leak location). With multiple pressure sensors, one should not only increase the likelihood of detecting a leak, but be able to deduce the general location of where the leak has occurred, and thus follow-up with geochemical sampling to determine the degree of the leak (Nogues et al., 2011).

The usefulness of chemical signals may improve over the long-term. It was observed that for all observation distances, the TDS signal is highest at 48 years. This may indicate that the extent of the TDS and pH signal could increase with longer observation times. However, for 48 years, the TDS changes were only seen in <20% and <10% of the samples at 100 m for the simulations with the top CO₂ and the top brine leakages, respectively. Simulations should be run longer to be able to comprehensively evaluate pressure, TDS and pH as long-term indicators.

Another significant inference from this work is how heterogeneity will affect the possible signals. Recall that the horizontal and vertical correlation lengths are based on data from the

1 regional High Plains Aquifer (Carroll et al., 2014). The horizontal correlation lengths varied
2 from 200-2,500 m. Therefore, there are instances where the flow barriers extend across the entire
3 model domain creating confined volumes that preserve the pressure signals located radially from
4 the leak. Consequently, the most expansive and earliest responses (e.g., 2 years) are when the
5 observation well is radially located from the leak (Fig. 13). At 2 years, there was already 20%
6 chance of detection out to 400m. The flow barriers allowed for this overpressure, and thus strong
7 radial signal.

8 The flow barriers also affect the likelihood of detection vertically from the leak. The likelihoods
9 of detection were generally higher but oscillated depending on the vertical correlation length of
10 the hydrogeologic facies. Hence, placement of sensors should try to consider the local
11 heterogeneity to minimize the masking affect it may have on pressure which can otherwise give
12 the earliest indication of a CO₂/brine breach into the aquifer.

13 This study only considers direct point-source sampling, but a comprehensive monitoring and
14 verification plan may benefit from indirect, remote-sensing techniques for carbon sequestration
15 activities. Trainor-Guitton et al. (2013) present a methodology for assessing the reliability of
16 electrical resistivity data to diagnose leakage due to a combined CO₂ and brine leakage. Future
17 work would be to include remote-sensing modeling such as this but also use this dataset to
18 optimize sensor location given the possible locations of signal (Gastelum and Porter, 2014).

19

1 **5. ACKNOWLEDGMENTS**

2 This work is part of the National Risk Assessment Partnership (NRAP) which is supported by
3 the U.S. Department of Energy's (DOE's) Office of Fossil Energy's Cross-cutting Research
4 program.

5

6. REFERENCES

- Cahill, A.G., Jakobsen, R. 2013 Hydro-geochemical impact of CO₂ leakage from geological storage on shallow potable aquifers: A field scale pilot experiment, *Int. J. of Greenhouse Gas Control*. 19, 678-688.
- Carle, S.F., 1999, T-PROG: Transitional probability geostatistical software users' guide; Davis, CA, University of California.
- Carroll, S. a., Keating, E., Mansoor, K., Dai, Z., Sun, Y., Trainor-Guitton, W., Brown, C., and Bacon, D., 2014, Key Factors for Determining Groundwater Impacts Due to Leakage from Geologic Carbon Sequestration Reservoirs: *International Journal of Greenhouse Gas Control*, v. 29, p. 153–168, doi: 10.1016/j.ijggc.2014.07.007.
- Dafflon, B., Wu, Y., Hubbard, S.S., Birkholzer, J.T., Daley, T.M., Pugh, J.D., Peterson, J.E., and Trautz, R.C., 2013, Monitoring CO₂ intrusion and associated geochemical transformations in a shallow groundwater system using complex electrical methods: *Environmental Science and Technology*, v. 47, p. 314–321, doi: 10.1021/es301260e.
- Daley, T.M., Myer, L.R., Peterson, J.E., Majer, E.L., and Hoversten, G.M., 2008, Time-lapse crosswell seismic and VSP monitoring of injected CO₂ in a brine aquifer: *Environmental Geology*, v. 54, p. 1657–1665, doi: 10.1007/s00254-007-0943-z.
- Gal, F., Proust, E., Humez, P., Braibant, G., Brach, M., Koch, F., Widory, D., and Girard, J.-F., 2013, Inducing a CO₂ Leak into a Shallow Aquifer (CO₂FieldLab Eurogia+ Project): Monitoring the CO₂ Plume in Groundwaters: *Energy Procedia*, v. 37, p. 3583–3593, doi: 10.1016/j.egypro.2013.06.251.
- Gasperikova, E., and Hoversten, G.M., 2008, Gravity monitoring of movement during sequestration: Model studies: *Geophysics*, v. 73, p. WA105–WA112, doi: 10.1190/1.2985823.
- Gastelum, J. a., and Porter, E. a., 2014, Optimizing Monitoring Designs under Alternative Objectives: *Energy Procedia*, v. 63, p. 3516–3522, doi: 10.1016/j.egypro.2014.11.380.
- In-Situ, 2014, Comparing Absolute and Gauged Pressure Sensors: *Innovations in Water Monitoring*, p. 4.

- 1 Keating, E., Dai, Z., Dempsey, D., and Pawar, R., 2014, Effective detection of CO₂ leakage: a
2 comparison of groundwater sampling and pressure monitoring: *Energy Procedia*, v. 63, p.
3 4163–4171, doi: 10.1016/j.egypro.2014.11.448.
- 4 Leuning, R., Etheridge, D., Luhr, A., and Dunse, B., 2008, Atmospheric monitoring and
5 verification technologies for CO₂ geosequestration: *International Journal of Greenhouse*
6 *Gas Control*, v. 2, p. 401–414, doi: 10.1016/j.ijggc.2008.01.002.
- 7 Lu, C., Sun, Y., Buscheck, T.A., Hao, Y., Joshua A. White, and Chiaramonte, L., 2012,
8 Uncertainty quantification of CO₂ leakage through a fault with multiphase and
9 nonisothermal effects: *Greenhouse Gases: Science and Technology*, v. 2, p. 445–459.
- 10 Mito, S., and Xue, Z., 2011, Post-Injection monitoring of stored CO₂ at the Nagaoka pilot site: 5
11 years time-lapse well logging results: *Energy Procedia*, v. 4, p. 3284–3289, doi:
12 10.1016/j.egypro.2011.02.248.
- 13 Nitao, J.J., 1998, User's manual for the USNT module of the NUFT code, version 2 (NP-phase,
14 NC-component, thermal): Lawrence Livermore National Laboratory.
- 15 Nogues, J.P., Nordbotten, J.M., and Celia, M. a., 2011, Detecting leakage of brine or CO₂
16 through abandoned wells in a geological sequestration operation using pressure monitoring
17 wells: *Energy Procedia*, v. 4, p. 3620–3627, doi: 10.1016/j.egypro.2011.02.292.
- 18 Ramirez, a. L., 2005, Stochastic inversion of electrical resistivity changes using a Markov Chain
19 Monte Carlo approach: *Journal of Geophysical Research*, v. 110, p. 1–18, doi:
20 10.1029/2004JB003449.
- 21 Schaap, M.G., Leij, F.J., and Van Genuchten, M.T., 2001, Rosetta: A computer program for
22 estimating soil hydraulic parameters with hierarchical pedotransfer functions: *Journal of*
23 *Hydrology*, v. 251, p. 163–176, doi: 10.1016/S0022-1694(01)00466-8.
- 24 Schmidt-Hattenberger, C., Bergmann, P., Kießling, D., Krüger, K., Rücker, C., Schütt, H., and
25 Group, K., 2011, Application of a Vertical Electrical Resistivity Array (VERA) for
26 monitoring CO₂ migration at the Ketzin site: First performance evaluation: *Energy*
27 *Procedia*, v. 4, p. 3363–3370, doi: 10.1016/j.egypro.2011.02.258.

- 1 Sun, A.Y., Zeidouni, M., Nicot, J.P., Lu, Z., and Zhang, D., 2013, Assessing leakage
2 detectability at geologic CO₂ sequestration sites using the probabilistic collocation method:
3 *Advances in Water Resources*, v. 56, p. 49–60, doi: 10.1016/j.advwatres.2012.11.017.
- 4 Trainor-Guitton, W.J., Ramirez, A., Yang, X., Mansoor, K., Sun, Y., and Carroll, S., 2013, Value
5 of information methodology for assessing the ability of electrical resistivity to detect
6 CO₂/brine leakage into a shallow aquifer: *International Journal of Greenhouse Gas Control*,
7 v. 18, p. 101–113, doi: 10.1016/j.ijggc.2013.06.018.
- 8 Vasco, D.W., Rucci, a., Ferretti, a., Novali, F., Bissell, R.C., Ringrose, P.S., Mathieson, a. S.,
9 and Wright, I.W., 2010, Satellite-based measurements of surface deformation reveal fluid
10 flow associated with the geological storage of carbon dioxide: *Geophysical Research*
11 *Letters*, v. 37, p. 1–5, doi: 10.1029/2009GL041544.
- 12 Wainwright, H.M., Finsterle, S., Zhou, Q., and Birkholzer, J.T., 2013, Modeling the performance
13 of large-scale CO₂ storage systems: A comparison of different sensitivity analysis methods:
14 *International Journal of Greenhouse Gas Control*, v. 17, p. 189–205, doi:
15 10.1016/j.ijggc.2013.05.007.
- 16 Wang, Z., and Small, M.J., 2014, A Bayesian approach to CO₂ leakage detection at saline
17 sequestration sites using pressure measurements: *International Journal of Greenhouse Gas*
18 *Control*, v. 30, p. 188–196, doi: 10.1016/j.ijggc.2014.09.011.
- 19 White, D.J., and Johnson, J.W., 2009, Integrated geophysical and geochemical research
20 programs of the IEA GHG Weyburn-Midale CO₂ monitoring and storage project: *Energy*
21 *Procedia*, v. 1, p. 2349–2356, doi: 10.1016/j.egypro.2009.01.305.
- 22 Yang, X., Chen, X., Carrigan, C.R., and Ramirez, A.L., 2014, Uncertainty quantification of CO₂
23 saturation estimated from electrical resistance tomography data at the Cranfield site:
24 *International Journal of Greenhouse Gas Control*, v. 27, p. 59–68, doi:
25 10.1016/j.ijggc.2014.05.006.