



GHGT-12

## Reduced Order Models for Prediction of Groundwater Quality Impacts from CO<sub>2</sub> and Brine Leakage

Liange Zheng<sup>a\*</sup>, Susan Carroll<sup>b</sup>, Marco Bianchi<sup>a</sup>, Kayyum Mansoor<sup>b</sup>, Yunwei Sun<sup>b</sup>, and Jens Birkholzer<sup>a</sup>

<sup>a</sup>Earth Sciences Division, Lawrence Berkeley National Laboratory,  
1 Cyclotron Road Berkeley, CA 94720

<sup>b</sup>Atmospheric, Earth and Energy Division, Lawrence Livermore National Laboratory 7000 East Ave., Livermore, CA 94551

---

### Abstract

A careful assessment of the risk associated with geologic CO<sub>2</sub> storage is critical to the deployment of large-scale storage projects. A potential risk is the deterioration of groundwater quality caused by the leakage of CO<sub>2</sub> and brine leakage from deep subsurface reservoirs. In probabilistic risk assessment studies, numerical modeling is the primary tool employed to assess risk. However, the application of traditional numerical models to fully evaluate the impact of CO<sub>2</sub> leakage on groundwater can be computationally complex, demanding large processing times and resources, and involving large uncertainties. As an alternative, reduced order models (ROMs) can be used as highly efficient surrogates for the complex process-based numerical models.

In this study, we represent the complex hydrogeological and geochemical conditions in a heterogeneous aquifer and subsequent risk by developing and using two separate ROMs. The first ROM is derived from a model that accounts for the heterogeneous flow and transport conditions in the presence of complex leakage functions for CO<sub>2</sub> and brine. The second ROM is obtained from models that feature similar, but simplified flow and transport conditions, and allow for a more complex representation of all relevant geochemical reactions. To quantify possible impacts to groundwater aquifers, the basic risk metric is taken as the aquifer volume in which the water quality of the aquifer may be affected by an underlying CO<sub>2</sub> storage project. The integration of the two ROMs provides an estimate of the impacted aquifer volume taking into account uncertainties in flow, transport and chemical conditions. These two ROMs can be linked in a comprehensive system level model for quantitative risk assessment of the deep storage reservoir, wellbore leakage, and shallow aquifer impacts to assess the collective risk of CO<sub>2</sub> storage projects.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of the Organizing Committee of GHGT-12

**Keywords:** groundwater; CO<sub>2</sub> leakage; brine leakage; reduced order models;

---

## 1. Introduction

Carbon dioxide (CO<sub>2</sub>) geologic storage is being considered as a possible measure to curb the anthropogenic emissions of greenhouse gases. A careful assessment of the risk associated with CO<sub>2</sub> geologic storage is critical to deployment of large scale CO<sub>2</sub> geological storage. One of the potential risks is the impact of potential CO<sub>2</sub> leakage from deep subsurface reservoirs on overlying groundwater aquifers. The leakage of CO<sub>2</sub> could affect such aquifers by increasing the concentration of carbonic acid and causing pH to drop, by mobilizing trace elements through mineral dissolution, desorption, and/or exchange, and by increasing dissolved solids, trace metals, and organics directly from the leaking brine [1-8]. Accurate prediction of groundwater impacts is complicated by the uncertainty and variability in model input data, including key parameters that are needed to describe leakage sources and the properties of aquifer systems. In principle, the full range of potential groundwater impacts could be assessed through rigorous numerical modeling, but a complete treatment of uncertainty and variability would be computationally prohibitive for most operators and regulators.

The National Risk Assessment Partnership (NRAP) is developing a science-based toolset for the analysis of the potential risks associated with changes in groundwater chemistry from CO<sub>2</sub> injection. This quantification approach is based on simulating coupled physical and chemical processes to predict how the natural system behaves over time, and it includes uncertainty quantification explicitly. In order to address uncertainty probabilistically, NRAP is developing efficient, reduced-order models (ROMs) as part of its approach. These ROMs are built from detailed, physics-based process models to provide confidence in the predictions over a range of conditions. However, the ROMs are designed to reproduce accurately the predictions from the computationally intensive process models at a fraction of the computational time, thereby allowing the utilization of Monte Carlo methods to probe variability in key parameters.

The development of ROMs that describe impacts to protected groundwater should consider all relevant physical flow and chemical processes. Developing ROMs based on high-fidelity numerical models that incorporate all relevant chemical reactions and transport fields is very challenging and computationally demanding. We circumvent these difficulties by correcting output from a ROM (referred to as the hydrology ROM, including hydrological processes and only few chemical reactions) with chemical scaling functions derived from more complex process models that consider comprehensive chemical reactions.

The natural system studied here is an alluvium aquifer consisting of layers of relatively high-permeable sands interbedded with low-permeable clays based on the hydrostratigraphic structure of the High Plains aquifer. Input parameters considered include: aquifer heterogeneity, permeability, porosity, regional groundwater flow, injection period, wellbore remediation, CO<sub>2</sub>, TDS, trace metal leakage rates over time, and geochemical parameters. The hydrology ROM and the chemistry scaling function are specific to thresholds described in Section 2 that represent no net degradation to the groundwater quality. The hydrology ROM is derived from physics-based simulations accounting for heterogeneous transport conditions, calcite dissolution/precipitation, and mitigation for leakage through a single wellbore and is described in Section 3. The chemistry scaling function corrects the hydrology ROM output for relevant geochemical reactions and is described in Section 4. We discuss the impacts to groundwater chemistry in Section 5. And in Section 6 we summarize the findings from this study.

## 2. Impact thresholds

The basic risk metric used in this study to quantify possible impacts to groundwater aquifers is the aquifer volume in which the water quality of the aquifer may be affected by CO<sub>2</sub> and brine leakage, which is defined as the aquifer volume with concentration of chemical components such as pH, TDS, trace metals and organic higher (or less) than their respective threshold values. The impact thresholds defined for pH, TDS, trace metals, and selected organics in Table 1 represent concentrations above the background water chemistry. Each threshold was calculated as the 95%-confidence, 95%-coverage tolerance from data collected in a 2010 U.S. Geological Survey (USGS) groundwater survey of 30 wells within the High Plains aquifer from an area outside the modeled domain. This data set was chosen because spatial and temporal data were not available within the modeled domain. We consider benzene, naphthalene, and phenol as representative, respectively, of benzene-toluene-ethylbenzene-xylene (BTEX) volatile aromatic compounds typically found in petroleum, poly-aromatic hydrocarbons (PAH), and phenol organic

compounds that could be present in the leaking brine [9]. Table 1 also includes regulatory standards which consist of primary or secondary maximum contaminant levels designated by the U.S. EPA (2009). Primary drinking water standards are for trace metals, such as As, Ba, Cd, Cr, Cu, Pb, BTEX and PAHs (as Benzo(a)pyrene) organics among others, and are legally enforced for the protection of public health by limiting the levels of contaminants in drinking water. Secondary drinking water standards, which include standards for Fe, Mn, and Zn, are non-enforceable guidelines regulating contaminants that may cause cosmetic or aesthetic effects in drinking water. Currently, drinking water standards have not been established for phenols as a group.

Table 1: Initial aquifer composition used in the simulations, with estimated mean concentration values and no-impact thresholds.

Parameter	Initial Value	Mean of Selected and Adjusted 2010 Data <sup>b</sup>	Impact Threshold <sup>c</sup>	U.S. EPA Regulatory Standard
pH	7.6 <sup>a</sup>	7.5 <sup>c</sup>	7.0	6.5
TDS	570 mg/L <sup>a,d</sup>	440 mg/L <sup>d</sup>	1,300 mg/L <sup>d,e</sup>	500 mg/L <sup>e</sup>
Arsenic	1.500 µg/L	1.500 µg/L	9.300 µg/L	10.0 µg/L
Barium	43.000 µg/L <sup>b</sup>	43.000 µg/L <sup>b</sup>	140.000 µg/L	2,000 µg/L
Cadmium	0.059 µg/L	0.059 µg/L	0.250 µg/L	5 µg/L
Chromium	1.000 µg/L	1.000 µg/L	3.900 µg/L	100 µg/L
Iron	5.400 µg/L <sup>b</sup>	5.400 µg/L <sup>b</sup>	43.000 µg/L <sup>b</sup>	300 µg/L
Lead	0.086 µg/L	0.086 µg/L	0.630 µg/L	15 µg/L
Manganese	0.350 µg/L <sup>d</sup>	0.350 µg/L <sup>d</sup>	7.000 µg/L <sup>d</sup>	50 µg/L
Benzene	0	<0.030 µg/L <sup>d</sup>	0.030 µg/L <sup>g</sup>	5 µg/L
Naphthalene	0	<0.200 µg/L <sup>d</sup>	0.200 µg/L <sup>g</sup>	700 µg/L
Phenol	0	<0.003 µg/L <sup>f</sup>	0.003 µg/L <sup>g</sup>	10,000 µg/L <sup>h</sup>

(a) Based on Carroll et al. [10].

(b) Geometric mean except for pH.

(c) 95%-confidence, 95%-coverage tolerance limit based on log values except for pH.

(d) Rounded to two significant digits.

(e) Threshold value exceeds regulatory standard; using the regulatory standard may result in widespread false positives under field conditions.

(f) As 4-Chloro-2-methylphenol

(g) Detection limit for the 2010 U.S. Geologic Survey National Water-Quality Assessment Program (NAWQA) sample data.

(h) Recommended Water Quality Criteria for Human Health, consumption of Water + Organism (74 FR 27535);

<http://water.epa.gov/scitech/swguidance/standards/criteria/current/index.cfm#hhtable>

### 3. Development of the hydrology reduced-order model

The hydrology ROM considered the variability of: wellbore leakage rates, mitigation times and arsenic, barium, cadmium, chromium, iron, lead, manganese, benzene, naphthalene, and phenol concentrations in the leaking brine. It also considered the uncertainty in the hydrologic properties of the aquifer. In this section, we describe the reactive transport models, parameter variability, and leakage profiles used to build this ROM, as well as the fidelity of the resulting threshold-specific ROMs. We used PSUADE, an uncertainty quantification code [11] to establish sampling points for the reactive transport simulations, to conduct parameter sensitivity analysis, and to develop ROMs from the simulated results.

#### 3.1. Reactive transport model

The first step in developing the hydrology ROM was to build statistical models of the aquifer that account for the variability and uncertainty in the lithology (as expressed by permeability). The natural system studied here is an alluvium aquifer consisting of layers of permeable sand interbedded with layers of clay with low permeability. The distribution of these sediments in the modeled domain was based on published data for the High Plains aquifer. We

used material-volume fraction and correlation lengths derived from well logs to build 1,000 geostatistical models that contain 3-D distributions of high and low permeability zones for use in the reactive-transport simulations. Each simulation uses a different geostatistical realization combined with other variable inputs to assess the impact of leakage on groundwater quality.

The numerical model domain extends to  $10,000 \text{ m} \times 5,000 \text{ m} \times 240 \text{ m}$  ( $x, y, z$ ) with one leakage source placed at a depth of 198 m. The regional groundwater flow is maintained by a 0.3% hydraulic gradient. This model employs a variably spaced mesh with the smallest elements located at the leakage source with a total of 190,350 nodes.

$\text{CO}_2$  carbonation and the dissolution of calcite are the only reactions included in the model, which account for the predominant reactions affecting pH. But dissolved sodium, chloride, trace metals and organic compounds are treated as conservative tracers in the model.

Conceptually,  $\text{CO}_2$  fluxes increase to a plateau during the onset of  $\text{CO}_2$  injection into the storage reservoir, maintain a constant flux during injection, and then decrease overtime to some minimal level (Fig. 1). The variability in the  $\text{CO}_2$  leakage profile was generated using four parameters:  $q_{\text{CO}_2}$ , 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  $\text{CO}_2$ . Brine leakage was characterized with a maximum and constant flux during injection, which falls off to a final flux after injection stops. Uncertainty in the brine leakage profile was generated using four parameters:  $q_{\text{BRN}}$ , the initial and maximum flux;  $\lambda q_{\text{BRN}}$ , the final flux;  $T_{1b}$ , the injection time;  $dT_{2b}$ , the duration of the transition between the maximum and final flux. An additional parameter,  $T_m$ , was included to represent wellbore mitigation time. Proposed parameter ranges for generalized  $\text{CO}_2$  and brine leakage models are given in Carroll et al.[12]

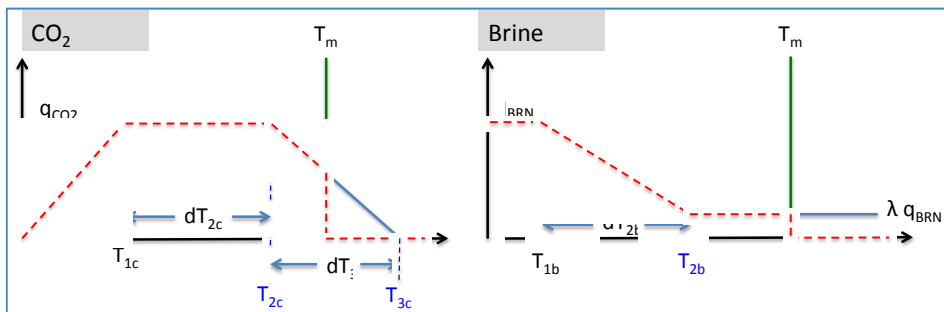


Fig. 1. Schematic of the  $\text{CO}_2$  and brine leakage model parameters and profiles in the generalized model.

### 3.2. Hydrology reduced order model

1000 simulations of the model described in Section 3.1 were conducted and the volumes of aquifer with concentration less than thresholds were computed. The emulations were carried out to reproduce the simulated results for specific thresholds for pH, TDS, As, Ba, Cd, Cr, Fe, Mn, Pb, benzene, naphthalene, and phenol using global and time-based correlation coefficients. The global correlations directly compare ROM and simulation output of all output times. Globally, the hydrology ROMs for pH, trace metals, and organics are robust and yield  $R^2$  between 0.9 and 0.99. This is not the case for TDS with global  $R^2 \approx 0.74$ . There is strong correlation between  $\text{CO}_2$  mass and pH plume volume, with a secondary dependence on mitigation time, which also correlates with cumulative mass of the leak. Correlations for the TDS, trace metal, and organic plume volumes on brine mass are more variable than pH volume because they are also dependent on brine concentration.

## 4. Development of the chemical scaling function

The chemical scaling function is used together with the hydrology ROM to estimate impact of leakage for trace metals and organics in the protected groundwater. Scaling functions are derived by comparing the output of models of deferent level of fidelity and are used to better emulate the response of an original high-fidelity model. The

approach is known as “variable-fidelity” or “multi-fidelity” modeling. Scaling functions can be multiplicative or additive depending on if the ratio rather than the difference between the responses from the lower and high-fidelity models is considered. In this work we found that the multiplicative approach provided the best accuracy when the scaling function  $g(X)$  was defined as:

$$g(X) = \frac{V_{HFM}(X)}{V_{LFM}(X^*)} \quad (1)$$

where  $V_{HFM}$  is the output volume calculated with the high-fidelity model with  $X$  input parameters and  $V_{LFM}$  is the corresponding volume obtained with a low fidelity model with  $X^*$  input parameters. To ensure the correspondence between the two models outputs,  $V_{LFM}$  and  $V_{HFM}$ ,  $X^*$  must be a subset of  $X$ . The exact form of the scaling function  $g(X)$  is typically unknown, especially for complex responses that depend on several physical and chemical processes such as those considered in this work. However, if we can find an approximation of the scaling function based on a relatively limited number of runs of the high- and low-fidelity models,  $V_{HFM}$  can then be approximated as:

$$\tilde{V}(X) = V_{LFM}(X^*) \cdot \tilde{g}(X) \quad (2)$$

Where  $\tilde{V}(X)$  is the emulated volume, and  $\tilde{g}(X)$  is the approximated scaling function designed to correct the value  $V_{LFM}$ . In this work,  $V_{LFM}(X^*)$  is the output from a simulation that does not consider geochemistry (i.e., the hydrology ROM) and  $\tilde{V}(X)$  is the corresponding output when all the relevant geochemical processes are taken into account. To estimate the latter, we performed reactive transport simulations that consider comprehensive chemical reactions as described below. The process of developing the approximate scaling function is analogous to the development of a traditional response surface ROM. However, scaling functions are generally less complex than typical response surface ROMs because their only purpose is to scale output derived from a similar model [13].

We developed scaling functions for As, Ba, Cd, Pb, benzene, naphthalene, and phenol using the thresholds listed in Table 1. Scaling functions for pH and TDS were not necessary because carbonate geochemistry, included in the hydrology ROM, accounts for the dominate changes in pH; and because mineral dissolution does not significantly alter the TDS estimated from brine leakage. To generate a scaling function for each of these outputs we followed a two-step procedure. In the first step we designed a numerical experiment to perform multiple runs of a model with no chemistry (so-called low-fidelity model, and is similar to the physics-based process models that are used to generate hydrology ROM) and a model with chemistry (so-called high-fidelity model) with the hydrology and geochemical input. In step two, we estimated an approximation of the true scaling function (Eq.1) for each of the considered outputs. These approximated scaling functions  $\tilde{g}(X)$  are polynomial. Its outputs are used to correct the output from the model without chemistry (Eq.2). In this step, we applied least-square fitting to calculate the coefficients of the polynomials representing the scaling functions. In general, a third-order polynomial provided the best match between the numerical and emulated ratios for metals, while second-order polynomial provided more accurate fitting for organic compounds. Note that scaling functions could take forms other than polynomial, but we use polynomial because it is simple and accurate enough.

The model with chemistry (so-called high-fidelity model) needed in the first step has the following features. It uses a single geostatistical realization of the aquifer lithology originally derived for the hydrology simulations (see Section 3.1) and a single CO<sub>2</sub> and brine leakage profile. Reactive-transport simulations were conducted with TOUGHREACT [14] and used the same domain size, boundary conditions, and hydrological gradient as described in Section 3.1 with a refined mesh near the leakage source and a courser mesh in the far field. Leakage occurred at a single point (x = 2,000 m, y = 2,500 m, z = 250 m). The geochemical model includes over 70 aqueous species, dissolution and precipitation kinetics for 14 minerals, 5 cation exchange reactions, 36 sorption reactions, and oxidation kinetics for 3 organics. Kinetic constants, mineral, cation exchange, and sorption mass balance reactions and constants are detailed in Carroll et al. [12]. Equilibrium constants for aqueous complexes and minerals are taken primarily from the THERMODDEM database [15] which can be downloaded from the website: <http://THERMODDEM.brgm.fr/index.asp>. Cation exchange reactions are included for Na, Ca, Mg, and Ba, as are

surface complexation reactions for As, Cd and Pb on goethite, illite, kaolinite and montmorillonite. Kinetic rate parameters for most rock-forming minerals were taken from Palandri and Kharaka [16], which are based mainly on experimental studies conducted under far-from-equilibrium conditions.

500 runs of models with no chemistry and models with chemistry were conducted and the corresponding  $V_{TM-HFM}$  (volume of aquifer with trace metal concentration higher than thresholds obtained in the high-fidelity model) to  $V_{pH-LFM}$  (volume of aquifer with pH less than threshold obtained in the low-fidelity model) for each pair of runs were obtained. Scaling functions for trace metals were developed from the ratios of  $V_{TM-HFM}$  to  $V_{pH-LFM}$  because of observed correlations between trace metal and pH plume volumes. Correlations were observed not only for As, Cd and Pb whose uptake reactions are known to be pH-dependent, but also for Ba, whose release is driven by Ca exchange at mineral surfaces [7]. The indirect correlation for Ba with pH is caused by the pH-dependence of calcite dissolution.

Scaling functions for organic compounds were developed from the ratios of  $V_{org-HFM}$  (volume of aquifer with organics concentration higher than thresholds obtained in the high-fidelity model) to  $V_{org-LFM}$  (volume of aquifer with organics concentration less than threshold obtained in the low-fidelity model) because the brine is the only source of organic compounds, and adsorption and oxidation reactions occur within the area where brine moves.

Goodness of fit was evaluated by comparing scaling factors estimated from the numerical simulations (Eq. 1) and the emulations (Eq. 2). Correlation coefficients range between  $R^2 = 0.76$  to  $0.86$  for trace metals and  $R^2 = 0.52$  to  $0.66$  for the organics. Improvements in the accuracy of the scaling functions may be achieved by using different mathematical forms for the scaling functions.

## 5. Results from the integrated ROM for groundwater chemistry

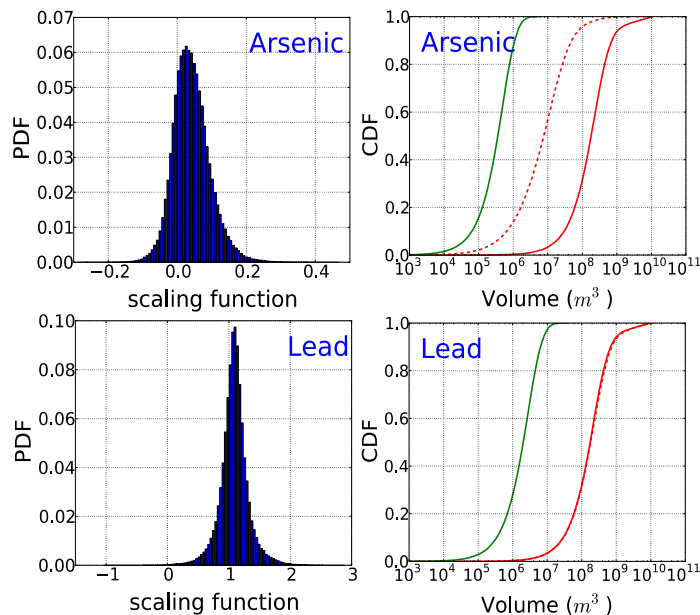


Fig. 2. Application of the scaling factors for arsenic and lead (left graphs) to the hydrology ROM (right graphs) where the solid green line represents the trace-metal plume volume computed with the hydrology ROM (as metals transported tracers without chemical reactions), the dashed red line represents the plume computed with the hydrology ROM corrected for geochemical reactions, and the solid red line represents the pH plume volume.

Fig. 2 and Fig. 3 compare plume volumes estimated from the hydrology ROM, which considered the reservoir source term for trace metals and organics, with plume volumes estimated with the scaling function that accounts for

the full suite of geochemical reactions. As mentioned above, scaling functions were not applied to pH and TDS plume volumes. Recall that the hydrology ROM considers variability and uncertainty in the geostatistical model and in CO<sub>2</sub> and brine leakage rates. Inclusion of trace metal chemistry using the scaling function tends to increase plume volumes by about 10 times for As and Ba and 100 times for Cd and Pb, as exemplified by the green and dashed-red cumulative distribution profiles for As and Pb in Fig. 2. The results suggest that a release of As, Ba, Cd, and Pb from the shallow aquifer sediments as the pH plume advances has a greater footprint when chemical reactions are taken into account than when trace metals are transported from leakage point as passive tracers. There are three possible outcomes from this analysis. One is that trace metal leakage sources need not be considered for risk assessments, because their input will be small relative to geochemical reactions within the protected aquifer. The second outcome is that the pH plume volumes derived from the hydrology ROM could be used as proxies for Ba, Cd, and Pb, because their plume volumes are about the same as the pH plume volume, as can be seen by comparing the solid and dashed red lines in Fig. 2. This is not the case for As, whose plume was about 10 times smaller than the pH plume. The third outcome is that direct assessment of trace metal release from the aquifer sediment could lower the uncertainty. This can easily be achieved through experiment [5]. It is important to conduct the experiment because of the complexity of the geochemical reactions and uncertainty associated with assessable surface area available for reaction. Direct knowledge of the bulk sediment response to CO<sub>2</sub> saturated waters might allow the trace metal thresholds to be directly correlated to variable pH thresholds.

In contrast to the trace metals, application of the scaling factor (see Eq. 2) yields large decreases in the organic plume volumes that are attributed to strong bio-degradation. Plume sizes are reduced by 10 times for benzene (Fig. 3), 100 times for phenol, and 1,000 times for naphthalene. Strong degradation of organics suggests minimal long-term impact of organic constituents.

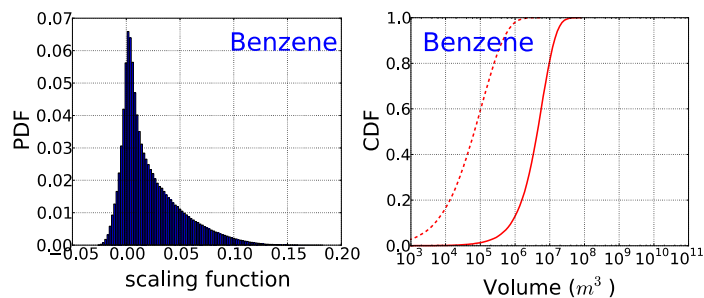


Fig. 3. Application of the scaling factors for benzene (left graph) to the hydrology ROM (right graph) where the solid and dashed lines represent the uncorrected and corrected plume volumes for benzene.

#### 4. Conclusion

We developed ROMs that describe changes in dilute groundwater chemistry if CO<sub>2</sub> and brine were to leak into an overlying an alluvium aquifer similar to the High Plains aquifer, Haskell County, Kansas, USA. The protocol allows uncertainty and variability in aquifer heterogeneity, fluid transport, and geochemical reactions to be collectively evaluated to assess potential changes in groundwater pH, TDS, As, Ba, Cd, Pb, benzene, naphthalene and phenol concentrations by developing a scaling function that can be applied to correct the output from hydrology ROM. The chemical scaling function is developed by comparing output from 500 reactive-transport simulations with and without chemistry for simplified 3-D models based on a single representation of aquifer heterogeneity and CO<sub>2</sub> and brine leakage profiles. No correction is needed for pH because calcite dissolution, the most important pH buffering processes, is included in the hydrology ROM. No correction is applied to TDS because calcite dissolution does not change the impacted volume above those predicted by the hydrology ROM. Adsorption and cation exchange are important for trace metal, but have very limited effect on TDS. Inclusion of chemical correction increases trace metal plumes by 10 to 100 times, suggesting that CO<sub>2</sub> leakage leaches trace metals from the aquifer sediments and should be considered in risk assessments. Although more study is needed, as are corrections for other trace metals, such as chromium, iron, manganese, and zinc, the strong correlation of trace metal and pH plume volumes suggests

that trace metal impact may be approximated by pH plumes, greatly simplifying the calculations. In contrast to the observed increases in trace metal plume volumes, inclusion of bio-degradation greatly reduces plume volumes for organics.

Although we have established a general protocol for developing a ROM that can be used to assess groundwater impacts by including physical and chemical uncertainty and variability, there is room for improvement. The model accuracy is not very high when emulations are compared to simulated output. Correlation coefficients range between  $R^2 = 0.75$  to  $0.85$  for trace metals and about  $0.60$  for the organics. Improvements may be achieved by increasing the number of simulations to more fully sample the parameter space and by applying different mathematical forms for the chemical scaling function.

In order to make ROMs applicable for other alluvium aquifers, we recommend that ROMs be developed for variable thresholds in light of the “no net degradation” requirement in the U.S. EPA guidelines (2010), because other aquifers are likely to have different background water chemistry. The full chemical and hydrology ROMs developed here are specific to the output thresholds that define no net degradation to the High Plains aquifer. Variable threshold ROMs could be extracted from the same data set.

## Acknowledgements

This work was funded by the Assistant Secretary for Fossil Energy, National Energy Technology Laboratory, National Risk Assessment Partnership, of the US Department of Energy under Contract No. DEAC02-05CH11231. We also thank Nicolas Spycher from Lawrence Berkeley National Laboratory for the help in the model development and review of this paper.

## References

- [1] Aiuppa A, Federico C, Allard P., Gurrieri S., Valenza M. Trace metal modeling of groundwater-gas-rock interactions in a volcanic aquifer: Mount Vesuvius, Southern Italy. *Chemical Geology* 2005; 216, 289–311.
- [2] Zheng L., Apps JA., Zhang Y., Xu T., Birkholzer JT. On mobilization of lead and arsenic in groundwater in response to CO<sub>2</sub> leakage from deep geological storage. *Chemical geology* 2009; 268, 281–297.
- [3] Apps JA., Zheng L., Zhang Y., Xu T., Birkholzer J. T. Evaluation of potential changes in groundwater quality in response to CO<sub>2</sub> leakage from deep geologic storage. *Transp. Porous Med.* 2010; 82, 215–246.
- [4] Kharaka Y K., Thordsen J J., Kakouros E. et al. Changes in the chemistry of shallow groundwater related to the 2008 injection of CO<sub>2</sub> at the ZERT field site, Bozeman, Montana. *Environmental Earth Sciences* 2010; 60, 273-284.
- [5] Little MG., Jackson R.B. Potential Impacts of Leakage from Deep CO<sub>2</sub> Geosequestration on Overlying Freshwater Aquifers. *Environmental Science & Technology* 2010; 44, 9225–9232.
- [6] Wilkin RT., Digiulio DC. Geochemical Impacts to Groundwater from Geologic Carbon Sequestration: Controls on pH and Inorganic Carbon Concentrations from Reaction Path and Kinetic Modeling. *Environmental Science & Technology* 2010; 44, 4821–4827.
- [7] Zheng L., Apps JA., Spycher N., Birkholzer JT., Kharaka YK., Thordsen J., Beers SR., Herkelrath WN., Kakouros E., Trautz RC. Geochemical modeling of changes in shallow groundwater chemistry observed during the MSU-ZERT CO<sub>2</sub> injection experiment. *International Journal of Greenhouse Gas Control* 2012; 7, 202–217.
- [8] Trautz R. C., Pugh JD., Varadharajan C., Zheng L., Bianchi M., Nico PS., Spycher NF., Newell D. L., Esposito, R. A., Wu, Y., Dafflon, B., Hubbard, S. S., Birkholzer, J. T. Effect of Dissolved CO<sub>2</sub> on a Shallow Groundwater System: A Controlled Release Field Experiment. *Environ. Sci. Technol.* 2012; 47, 298-305.
- [9] Zheng L., Spycher N., Birkholzer J. T., Xu T., Apps J., Kharaka Y. Modeling studies on the transport of benzene and H<sub>2</sub>S in CO<sub>2</sub>-water systems, Technical Report: LBNL-4339E, Lawrence Berkeley National Laboratory: Berkeley, CA, 2010.
- [10] Carroll S., Hao Y., Aines R. Geochemical detection of CO<sub>2</sub> in dilute aquifers. *Geochemical Transactions* 2009; 10, 4.
- [11] Tong C. PSUADE User’s Manual, LLNL-SM-407882, Lawrence Livermore National Laboratory, Livermore, CA, 2005.
- [12] Carroll SA., Bianchi M., Mansoor K., Zheng L., Sun Y., Spycher N., Birkholzer J. Reduced-Order Model for Estimating Impacts from CO<sub>2</sub> Storage Leakage to Alluvium Aquifers: Third-Generation Combines Physical and Chemical Processes, NRAP-TRS-III-XXX-2014, NRAP Technical Report Series, U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2014; p 51.
- [13] Razavi S., Tolson B. A., Burn D. H. Review of surrogate modeling in water resources. *Water Resour. Res.* 2012; 48, W07401. DOI:10.1029/2011WR011527
- [14] Xu T., Sonnenthal E., Spycher N., Pruess K. TOUGHREACT: A Simulation Program for Non-isothermal Multiphase Reactive Geochemical Transport in Variably Saturated Geologic Media. *Computers and Geosciences* 2006; 32, 145–165.
- [15] Blanc, P., Lassin, A., Piantone, P., Azaroual, M., Jacquemet, N., Fabbri, A., Gaucher, E. C. Thermodem: A geochemical database focused on low temperature water/rock interactions and waste materials. *Applied Geochemistry* 2012, 27, 2107–2116.



- [16] Palandri J., Kharaka Y. K. A compilation of rate parameters of water-mineral interaction kinetics for application to geochemical modeling, USGS Open File report 2004-1068, U.S. Geological Survey, 2004; p 64.