

Reduced Order Models for Prediction of Groundwater Quality Impacts from CO₂ and Brine Leakage—Application to the High Plains Aquifer

September 2014

Office of Fossil Energy

NRAP-TRS-III-00X-2014

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference therein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed therein do not necessarily state or reflect those of the United States Government or any agency thereof.

This report has been reviewed by Yingqi Zhang and approved for public release.

Cover Illustration: Second order polynomial linking function for estimating the volume of pH < 6.5 after 180 days of leakage. Points represent the calculated responses from the simple and complex models.

Suggested Citation: Bianchi, M.; Zheng, L.; Spycher, N.; Birkholzer, J. *Reduced-Order Models for Prediction of Groundwater Quality Impacts from CO₂ and Brine Leakage—Application to the High Plains Aquifer*, NRAP-TRS-III-00X-2014; NRAP Technical Report Series; U.S. Department Energy, National Energy Technology Laboratory: Morgantown, WV, 2014; p XX.

An electronic version of this report can be found at: www.netl.doe.gov/nrap

Reduced Order Models for Prediction of Groundwater Quality Impacts from CO₂ and Brine Leakage—Application to the High Plains Aquifer

Marco Bianchi¹, Liange Zheng¹, Nicolas Spycher¹, and Jens Birkhozler¹

**¹Earth Sciences Division, Lawrence Berkeley National Laboratory,
1 Cyclotron Road, Berkeley, CA 94720**

NRAP-TRS-III-00X-2014

Level III Technical Report Series

X Month 2014

This page intentionally left blank.

Table of Contents

1. ABSTRACT OR EXECUTIVE SUMMARY.....	1
2. INTRODUCTION.....	2
3. GENERATION II ROM FOR REACTIVE TRANSPORT PROCESSES.....	4
3.1 THE BASE MODEL	4
3.2 GLOBAL SENSITIVITY ANALYSIS	12
3.3 ROM DEVELOPMENT	15
4. LINKING FUNCTION	19
4.1 INTRODUCTION AND BACKGROUND	19
4.2 THE LINKING FUNCTION METHOD.....	21
4.3 DEVELOPMENT OF LINKING FUNCTION	22
4.4 LINKING FUNCTIONS AND RESULTS	30
5. SUMMARY AND DISCUSSION	35
6. REFERENCES.....	36
APPENDIX A: LINKING FUNCTIONS PARAMETERS.....	40

List of Figures

Figure 1. The leakage rate of CO ₂ and brine as a function of time.....	4
Figure 2. Spatial distribution of total dissolved CO ₂ at 50, 100, 200 years.....	6
Figure 3. Spatial distribution of total dissolved CO ₂ at 50, 100, 200 years.....	7
Figure 4. Spatial distribution of sorbed Pb (as (mon_) ₂ Pb, a surface Pb species on smectite) at 50, 100, 200 years.....	7
Figure 5. Spatial distribution of dissolved Pb concentration at 50, 100, 200 years.....	8
Figure 6. Spatial distribution of dissolved Cd concentration at 50, 100, 200 years.	8
Figure 7. Spatial distribution of dissolved As concentration at 50, 100, 200 years.....	9
Figure 8. Spatial distribution of dissolved As concentration at 200 years in the base-case run and a sensitivity run that does not consider any chemical reactions.	10
Figure 9. Spatial distribution of dissolved As concentration at 200 years in the base-case run and a sensitivity run that inherits all the chemical reactions of the base-case run but with As concentration in the leaking brine at 10 ⁻⁵ mol/kg, about 1.5 order of magnitude higher than that in the base-case run.....	10
Figure 10. Spatial distribution of dissolved Cd concentration at 200 years in the base-case run and a sensitivity run that inherits all the chemical reactions in the base-case run but with Cd concentration in the leaking brine at 10 ⁻⁶ mol/kg, about 1.5 order of magnitude higher than that in the base run.	11
Figure 11. Spatial distribution of dissolved Pb concentration at 200 years in the base-case run and which inherits all the chemical reactions in the base-case run but with Pb concentration in the leaking brine at 10 ⁻⁵ mol/kg, about 1.75 order of magnitude higher than that in the base run.	11
Figure 12. The Sobol global sensitivity index for the volume of aquifer with pH<6.5.....	14
Figure 13. The Sobol global sensitivity index for the volume of aquifer with TDS>500 mg/L.....	14
Figure 14. The Sobol global sensitivity index for the volume of aquifer with concentration of As > MCL.....	15
Figure 15. The R ² of ROMs for the volume for aquifer with pH<6.5, TDS>500 mg/kg, and As concentration >MCL.....	16
Figure 16. Comparison of emulated volumes of aquifer with TDS>500 mg/L with ROM and simulated ones with numerical models for the 70 years.	16
Figure 17. Comparison of emulated volumes of aquifer with TDS>500 mg/L with ROM and simulated ones with numerical models for the 200 years.	17
Figure 18. Comparison of emulated volumes of aquifer with pH<6.5 with ROM and simulated ones with numerical models for the 70 years.	17

Figure 19. Comparison of emulated volumes of aquifer with As concentration >MCL with ROM and simulated ones with numerical models for the 70 years.	18
Figure 20. Heterogeneous distribution of two hydrostratigraphic units generated with T-PROGS (a). Numerical model mesh used for the TOUGHREACT simulations (b). The black circle indicates the location of the CO ₂ and brine leakage point.	24
Figure 21. Initial and boundary conditions for the simple 2-D model. The same applies for the complex model.	25
Figure 22. Maximum CO ₂ and brine leakage rates over time.....	26
Figure 23. Simple 2-D model base-case simulation results after 200 years of continuous leakage (unreactive transport). pH distribution (a), AsO ₃ concentration (b), Pb ²⁺ concentration (c), Cd ²⁺ concentration (d).	Error! Bookmark not defined.
Figure 24. Complex 2-D model base-case simulation results after 200 years of continuous leakage (reactive transport). pH distribution (a), AsO ₃ concentration (b), Pb ²⁺ concentration (c), Cd ²⁺ concentration (d).	Error! Bookmark not defined.
Figure 25. Second order polynomial linking function for estimating the volume of pH < 6.5 after 180 days of leakage. Points represent the calculated responses from the simple and complex models.	32
Figure 26. Comparison between complex model responses and those of the linking functions. The simulation time is 200 years. As (a); pH (b); TDS (c).....	33
Figure 27. Flow chart for applying the linking function approach.	34

List of Tables

Table 1. Initial chemical composition of groundwater in the model. Concentrations are in molal units (except for pH and Eh). Note that the concentration of Cl was somewhat modified from the original data to maintain charge balance. The Fe ⁺³ concentration is calculated by assuming equilibrium with goethite.	5
Table 2. Concentration of contaminants in the leaking brine (in molal units). The concentrations of other brine constituents are the same as in Table 1.	6
Table 3: Input variables and ranges for global sensitivity analysis	13
Table 5. Input parameters for the simple 2-D model base case	27
Table 6: Input parameters of the numerical experiment.	31

Acronyms, Abbreviations, and Symbols

Term	Description
EPA	U.S. Environmental Protection Agency
LANL	Los Alamos National Laboratory
LBNL	Lawrence Berkeley National Laboratory
LF	Linking function
LLNL	Lawrence Livermore National Laboratory
MCL	Maximum Contaminant Level
mon	Montmorillonite
mg/L	Milligrams/liter
NRAP	National Risk Assessment Partnership
PNNL	Pacific Northwest National Laboratory
PSUADE	Problem Solving environment for Uncertainty Analysis and Design Exploration software
ROM	Reduced-order model
SA	Sensitivity analysis
SCC	Supercritical CO ₂
TDS	Total dissolved solids
UA	Uncertainty assessment
UQ	Uncertainty quantification
WWC5	Water Well Completion Records (WWC5) Database

DRAFT

Acknowledgments

This work was completed as part of the National Risk Assessment Partnership (NRAP) project. Support for this project came from the Department of Energy's (DOE) Office of Fossil Energy's Crosscutting Research program. The authors wish to acknowledge Robert Romanosky (National Energy Technology Laboratory (NETL) Strategic Center for Coal) and Regis Conrad (DOE Office of Fossil Energy) for programmatic guidance, direction, and support.

NRAP funding was provided to Lawrence Berkeley National Laboratory (LBNL) under U.S. DOE Contract No. DE-AC02-05CH11231.

The authors also thank Yingqi Zhang and Daniel Hawkes for constructive reviews of this report.

DRAFT

EXECUTIVE SUMMARY

In this study, modeling was performed to represent the complex hydrogeological and geochemical conditions in a heterogeneous aquifer by using two separate ROMs. The first ROM is derived from a high-fidelity model that accounts for the heterogeneous flow and transport conditions in the presence of multiple leakage wells. This ROM, developed by Lawrence Livermore National Laboratory (LLNL) takes into account uncertainties related to flow, transport, and leakage parameters, but this ROM has a simplified representation of chemical reactions. The second ROM is obtained from models that feature greatly simplified flow and transport conditions, but allow for a more complex representation of all relevant geochemical reactions. This ROM, developed by Lawrence Berkeley National Laboratory (LBNL) takes into account uncertainties related to chemical parameters and reactions. Both ROMs specifically address the physical and chemical properties of the High Plains Aquifer (Becker et al., 2002).

Neither ROM can separately provide an accurate prediction of the risk profile volume, because of the simplifications inherent in these models. Accurate prediction could be achieved with a very complex 3-D reactive transport model this model would be too demanding computationally. Therefore, development of an alternative approach that allows linking of the two ROMs and, in particular, a correction of the risk profile volumes estimated by the two separate ROMs, was conducted.

The development of a linking function was accomplished by: (1) the establishment of two simple models— a 1-D model with homogenous flow and transport field and single leakage point, but including as many chemical reactions as possible—and a 2-D model considering aquifer heterogeneity but no reactions; (2) the development of a complex model that incorporates all the parameters and physical and chemical processes of both simple models; (3) multiple runs of both simple and complex models; and (4) estimation of the linking functions based on those runs.

This report describes the development and usage of the ROMs and linking function.

1. INTRODUCTION

CO₂ 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 CO₂ geologic storage is critical to deployment of large-scale CO₂ geological storage. One of the potential risks is the impact of potential CO₂ leakage from deep subsurface reservoirs on overlying groundwater aquifers. The leakage of CO₂ could affect such aquifers in several ways. First, the dissolution of CO₂ in water increases the concentration of dissolved carbonic acid, and thus the pH of the groundwater drops. Second, trace elements could be mobilized through mineral dissolution, desorption reactions, and/or exchange reactions involving H⁺ and other mobilized constituents (e.g., Aiuppa et al., 2005; Zheng et al., 2009; Kharaka et al., 2010; Little and Jackson, 2010; Wilkin and Digiulio, 2010; Zheng et al., 2012a; Trautz et al., 2012). Third, because supercritical CO₂ (SCC) is also an excellent solvent for organic compounds (Anitescu and Tavlarides, 2006; Kharaka et al., 2009), concerns have been raised about the potential mobilization of organic constituents from depth and subsequent transport to shallow drinking water bodies via leakage pathways (Zheng et al. 2010; 2013).

This particular risk is one of the major risks considered in risk profiles developed by the National Risk Assessment Partnership (NRAP). NRAP is a multi-laboratory research program aiming to provide the scientific underpinning for risk assessment with respect to the long-term storage of CO₂. Numerical models are the primary tools employed in the development of risk profiles. Because numerical models for evaluating the impact of CO₂ leakage on groundwater are very complicated and also involve large uncertainties, more efficient models are needed during the development of risk profiles. Reduced-order models (ROM) were therefore proposed to act as surrogates for the complex process-based numerical models.

So far, two generations of ROM have been developed for the NRAP project. The first-generation ROM uses pH and total dissolved solids (TDS) as the risk-monitoring metrics and focuses on the temporal and spatial evolution of pH and TDS in response to CO₂ and/or brine intrusion into a shallow aquifer. The second-generation ROM includes the temporal and spatial evolution of trace metals, such as lead (Pb), arsenic (As), and cadmium (Cd), that might be mobilized from the aquifer sediments as a result of leakage. This document describes the development of a second-generation ROMs to investigate the impact of brine and CO₂ leakage on groundwater in the High Plain Aquifer System, in fulfillment of NRAP milestones D2.3.5.3.

The development of ROMs generally relies on conducting a limited number of high-fidelity numerical simulations that consider all relevant flow, transport, and chemical processes that could potentially have an impact on CO₂ and brine leakage into groundwater. These high-fidelity simulations are then used to “train” simpler ROMs (e.g., look-up tables, functional relationships) that sufficiently represent their results for a wide range of uncertain input parameters. Developing such high-fidelity numerical models that incorporate 3-D heterogeneous flow, and transport fields, and all relevant chemical reactions is very challenging. For our purpose, with currently available codes, such a model would be too computationally demanding to make the derivation of ROMs practically feasible.

In this study, we make an attempt to represent the complex hydrogeological and geochemical conditions in a heterogeneous aquifer by using two separate ROMs. The first ROM is derived from a high-fidelity model that accounts for the heterogeneous flow and transport conditions in the presence of multiple leakage wells. This ROM, developed by Lawrence Livermore National

Laboratory (LLNL) and referred to as LLNL ROM, takes into account uncertainties related to flow, transport, and leakage parameters, but this ROM has a simplified representation of chemical reactions. The second ROM is obtained from models that feature greatly simplified flow and transport conditions, but allow for a more complex representation of all relevant geochemical reactions. This ROM, developed by Lawrence Berkeley National Laboratory (LBNL) and referred to as LBNL ROM, takes into account uncertainties related to chemical parameters and reactions. (See Section 3 for development of the LBNL ROM.) Both ROMs specifically address the physical and chemical properties of the High Plains Aquifer (Becker et al., 2002). A parallel effort is conducted for the Edwards Aquifer example by NRAP groups based at the Pacific Northwest National Laboratory (PNNL) and Los Alamos National Laboratory (LANL) (Bacon, 2013).

NRAP system-level model results describe the environmental risk of geological CO₂ storage for so-called risk metrics. Regarding possible impacts to groundwater aquifers, the currently defined risk metrics are the aquifer volume in which the water quality of the aquifer may have been affected by an underlying CO₂ storage projects. Examples of risk profiles include the volume in which pH is reduced and the volume in which arsenic concentrations are elevated above a threshold value. Clearly, neither the LLNL ROM nor the LBNL ROM can separately provide an accurate prediction of the risk profile volume, because of the simplifications inherent in these models. Accurate prediction could be achieved with a very complex 3-D reactive transport model, but, as mentioned above, this model would be too demanding computationally. Therefore, we seek to develop an alternative approach that allows linking of the two ROMs and, in particular, a correction of the risk profile volumes estimated by the two separate ROMs. This approach, defined here as the linking function method, is described in Section 4 of this report.

As discussed in Section 3, the reactive transport ROM is based on a 1-D reactive transport model of a representative flow path of the High Plains Aquifer. The model considers aqueous complexation, mineral dissolution/precipitation, and adsorption/desorption via surface complexation. This basic geochemical model draws on lessons learned from detailed mineralogical and experimental studies on groundwater sediments exposed to elevated levels of dissolved CO₂ (Zheng et al., 2012a; Trautz et al., 2012). The second-generation ROM was thereafter developed based on results from the 1-D flow-path model. Outputs from the ROM include the aquifer volume with pH reduced to below 6.5, the aquifer volume with total dissolved solid (TDS) elevated above 500 mg/L, the aquifer volume with concentration of As above the maximum contaminant level (MCL), the aquifer volume with concentration of Pb above the MCL, and the aquifer volume with concentration of Cd above the MCL. The inputs to the ROM are several parameters related to chemical reactions. In Section 4, we applied the linking function method to combine the LBNL ROM and the LLNL ROM, so that the uncertainties of all relevant physical and chemical parameters are taken into account in the estimation of the impacted aquifer volume. The linking function method is tested and verified by comparing the predictions of the developed functions against the output of a detailed high-fidelity model.

2. SECOND-GENERATION ROM FOR REACTIVE TRANSPORT PROCESSES

2.1 THE BASE MODEL

2.1.1 Model setup

The base 1-D model attempts to mimic a stream line in the 3-D flow field within the LLNL model of the High Plains Aquifer. The simulation domain is 10,000-m long and discretized into 1000 equidistant gridblocks 10-m wide and 10-m high. A hydraulic gradient of 0.003 is applied by fixing the pressure at gridblocks on the left and right boundaries. A single leaking point is assigned at $x = 1000$ m, with CO₂ and brine leakage rates shown in Figure 1. These rates represent a hypothetical leakage pathway related to a deep leaky well connecting a deep geologic reservoir for CO₂ storage with a shallow groundwater resource. The temporal evolution and magnitude of leakage is based on the first-generation ROM for wellbore leakage.

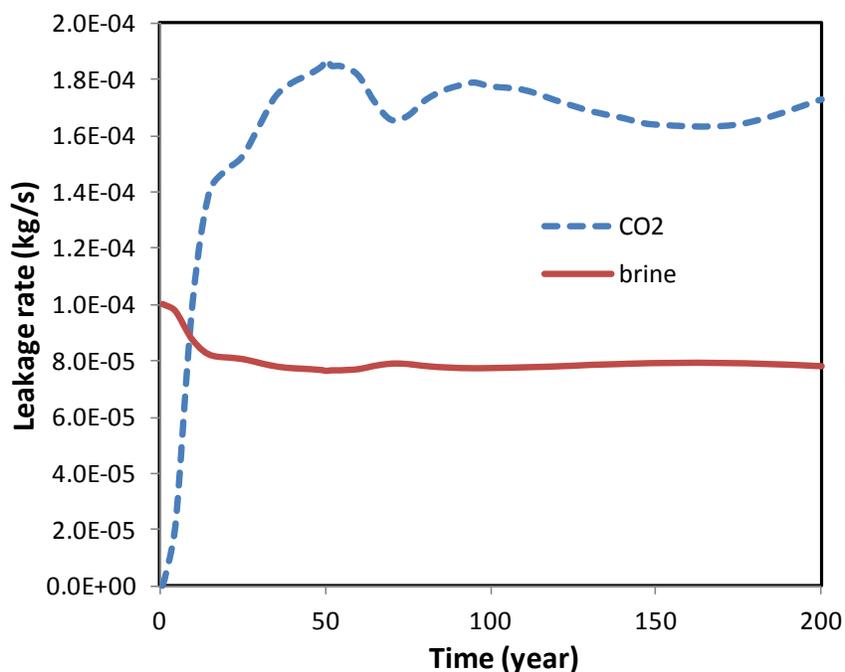


Figure 1. The leakage rate of CO₂ and brine as a function of time.

In the geochemical model, the chemical reactions considered are aqueous complexation, mineral dissolution/precipitation, cation exchange, and adsorption/desorption via surface complexation. Details of these reactions are given in NRAP deliverables 3.5.1d, 3.5.1e (Zheng et al., 2012b) and 3.5.1f (Zheng et al., 2012c). In the current model, the dissolution of calcite and surface protonation reactions are the main pH buffering processes. Surface complexation reactions on goethite, illite, kaolinite, and smectite are the dominant reactions that control the release of As, Pb and Cd.

In 1999, water samples from 74 randomly selected domestic water-supply wells completed in the Ogallala Formation of the central High Plains area were collected as part of the High Plains Regional Ground-Water Study conducted by the U.S. Geological Survey's National Water-Quality Assessment Program (Becker et al., 2002). The samples were analyzed for about 170 water-quality constituents

that included physical parameters, total dissolved solids and major ion concentrations, nutrients and dissolved organic carbon, trace elements, pesticides, volatile organic compounds, and radon. In 2010, 30 of 74 wells wereresampled. The initial concentration of major and trace elements in the model are the arithmetic average of concentration data from both 1999 and 2010, which are listed in Table 1.

Table 1. Initial chemical composition of groundwater in the model. Concentrations are in molal units (except for pH and Eh). Note that the concentration of Cl was somewhat modified from the original data to maintain charge balance. The Fe⁺³ concentration is calculated by assuming equilibrium with goethite.

Primary Species	Total Concentration (molal)
Ca ⁺²	1.37E-03
Mg ⁺²	8.74E-04
Na ⁺	1.60E-03
K ⁺	1.13E-04
Cl ⁻	1.85E-03
HS ⁻	6.48E-04
SO ₄ ⁻²	5.77E-05
HCO ₃ ⁻	3.95E-03
H ₄ SiO ₄ (aq)	5.33E-04
Fe ⁺²	1.01E-06
Fe ⁺³	8.97E-16
AlO ₂ ⁻	1.57E-07
Sb(OH) ₃ (aq)	5.59E-10
H ₃ AsO ₃ (aq)	2.33E-08
Ba ⁺²	8.10E-07
Cd ⁺²	6.91E-10
Pb ⁺²	1.57E-09
HSe ⁻	5.34E-08
Eh	-0.26 volts
pH	7.4

The concentration of Cl, Pb, Cd, and As in the leaking brine are four of the six variables for uncertainty quantifications, as described below. Their concentrations in the base case are listed in Table 2. The pH and concentration of dissolved CO₂ are obtained by saturating the initial water with CO₂ at the hydraulic pressure of the aquifer. For the rest of species in the leaking brine, their concentrations are assumed to be the same as in the initial water.

Table 2. Concentration of contaminants in the leaking brine (in molal units). The concentrations of other brine constituents are the same as in Table 1.

Primary Species	Total Concentration (molal)
pH	5.36
HCO ₃ ⁻	0.44
Na ⁺	1.00
Cl ⁻	1.00
H ₃ AsO ₃ (aq)	3.16E-07
Cd ⁺²	3.16E-08
Pb ⁺²	1.77E-07

2.1.2 Model results

Simulations were conducted with TOUGHREACT (Xu et al., 2012). The leakage of CO₂ into the aquifer leads to an increase in concentrations of dissolved CO₂ (Figure 2) and a subsequent drop in pH (Figure 3). The drop in pH induces desorption reactions, as illustrated by the concentration changes of montmorillonite 2Pb, a surface Pb species on smectite (Figure 4). Consequently, Pb is released into to groundwater and the concentration of Pb increases (see Figure 5). Similarly, the concentrations of Cd and As increase in groundwater, as shown in Figure 6 and 7.

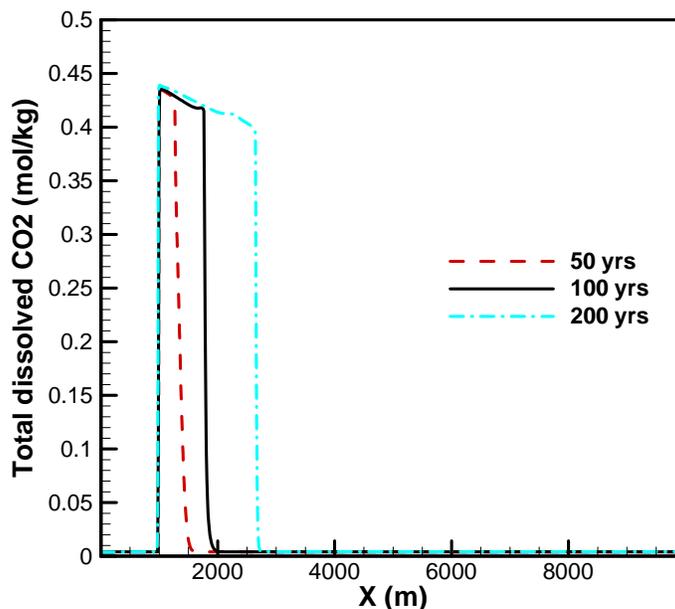


Figure 2. Spatial distribution of total dissolved CO₂ at 50, 100, 200 years.

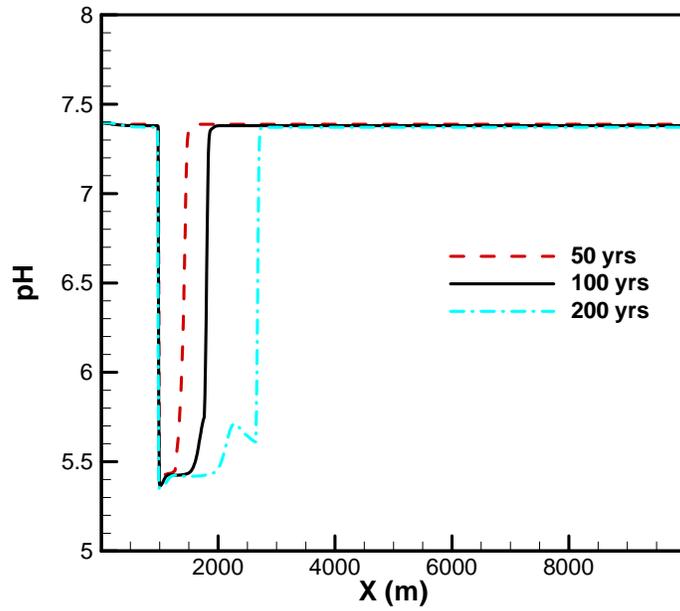


Figure 3. Spatial distribution of total dissolved CO₂ at 50, 100, 200 years.

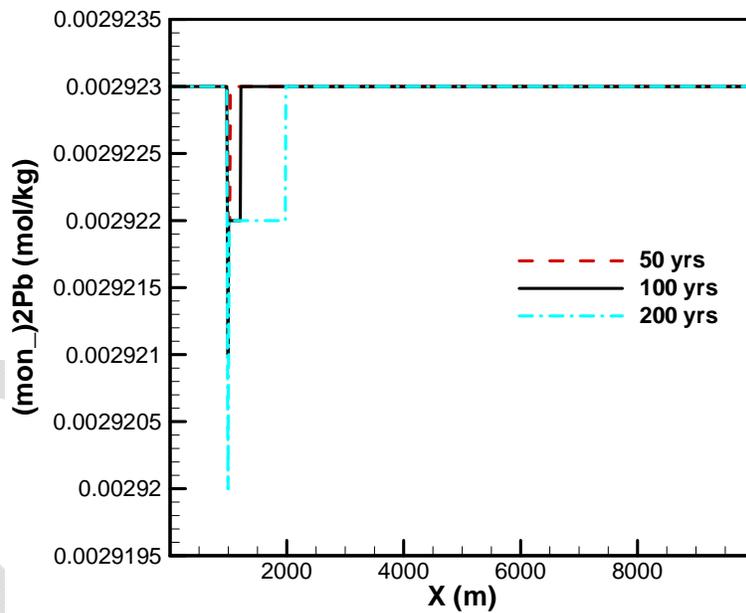


Figure 4. Spatial distribution of sorbed Pb (as (mon_)2Pb, a surface Pb species on smectite) at 50, 100, 200 years.

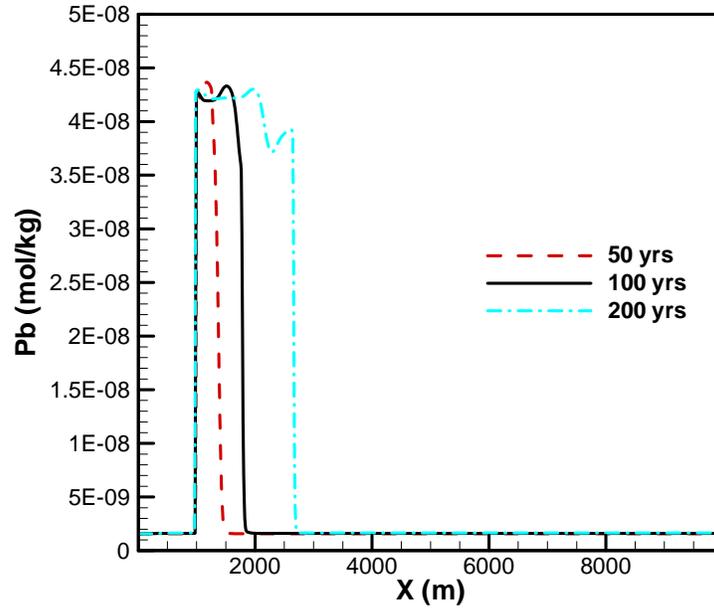


Figure 5. Spatial distribution of dissolved Pb concentration at 50, 100, 200 years.

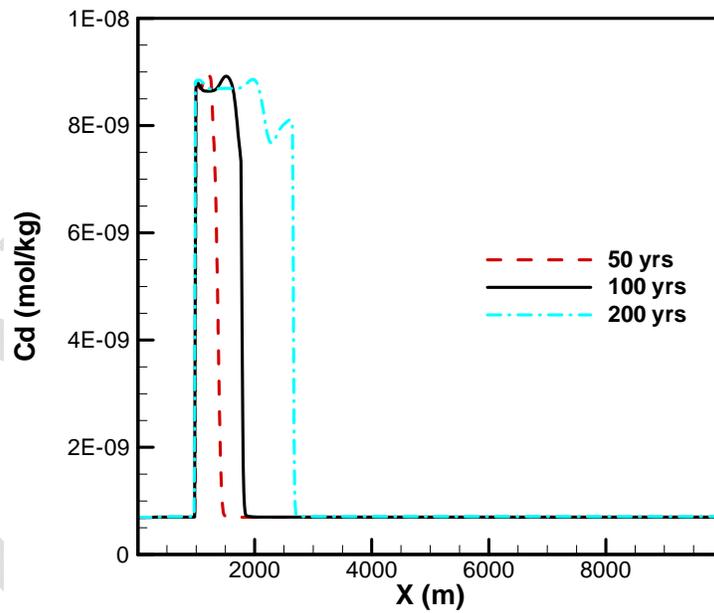


Figure 6. Spatial distribution of dissolved Cd concentration at 50, 100, 200 years.

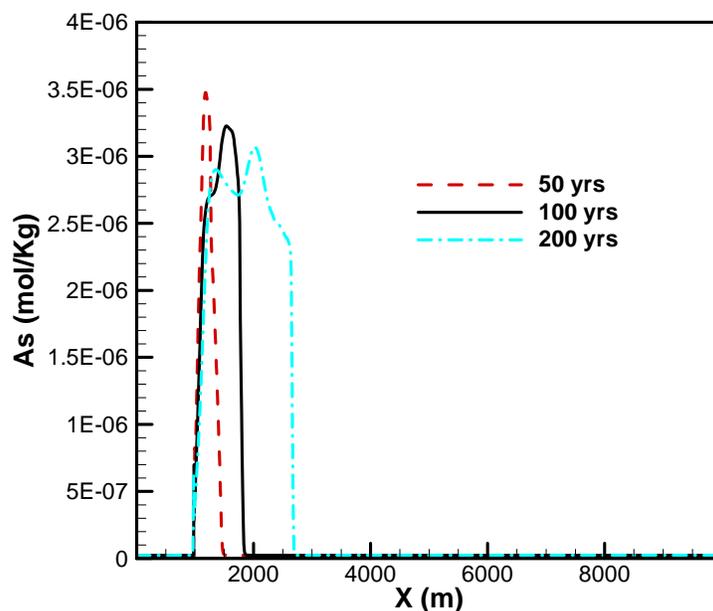


Figure 7. Spatial distribution of dissolved As concentration at 50, 100, 200 years.

The increase in trace metal concentrations within the aquifer is caused by two contamination sources: the leaking brine, which contains trace metals, and mobilization of trace metals within the aquifer itself via desorption from mineral surfaces. One interesting question is, which one of these two sources plays the major role? Figure 8 shows model results for As in a sensitivity run, which does not consider any reactions. In this run, concentration increase in the aquifer is controlled exclusively by the concentrations of trace metal in the leaking brine. The concentration of As predicted in this sensitivity run is very low in comparison with that in the base-case run (Figure 8). Another sensitivity run inherits all the chemical reactions of the base-case run, but with As concentration in the leaking brine as high as 10^{-5} . In this case, the predicted As concentration is still the same as in the base-case run (Figure 9). Similarly, as shown in Figures 10 and 11, increasing the concentration of Pb and Cd in the leaking brine does not make any difference to the concentration of Pb and Cd predicted in the aquifer. These results suggest that the leaking brine does not contribute significantly to the rise in concentrations of trace metals in the aquifer.

There are two reasons for this behavior. First, the leaking brine is significantly diluted by the upstream uncontaminated water. Second, the desorption/adsorption reactions act as a buffer. If the concentration of trace metals in the leaking brine is higher than in the aquifer, the desorption by pH decrease will be impeded to some extent. On the other hand, if the concentration of the trace metal in the brine is lower than that in the aquifer, the effect of pH on desorption will be increased. In either case, however, the ultimate concentration is controlled by sorption/desorption reactions rather than the concentration of leaking brine (because sorption sites on aquifer sediments are not fully saturated). Note that this observation is only valid for the scenario used in this report. If in some particular cases, the quantity of leaking brine is large enough and the concentration of trace metal is high enough to surpass the buffer by surface complexation (i.e., high enough to saturate sorption sites), the leaking brine could play a major role in determining the resulting trace metal concentration in the aquifer.

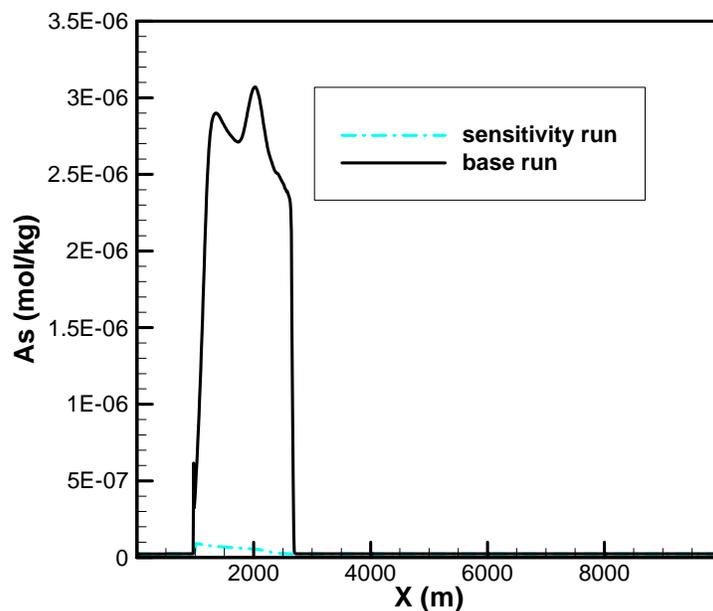


Figure 8. Spatial distribution of dissolved As concentration at 200 years in the base-case run and a sensitivity run that does not consider any chemical reactions.

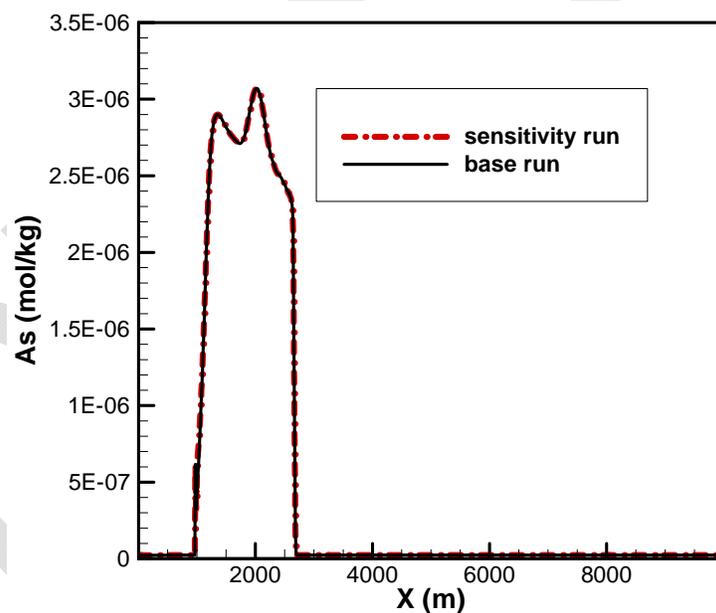


Figure 9. Spatial distribution of dissolved As concentration at 200 years in the base-case run and a sensitivity run that inherits all the chemical reactions of the base-case run but with As concentration in the leaking brine at 10^{-5} mol/kg, about 1.5 order of magnitude higher than that in the base-case run.

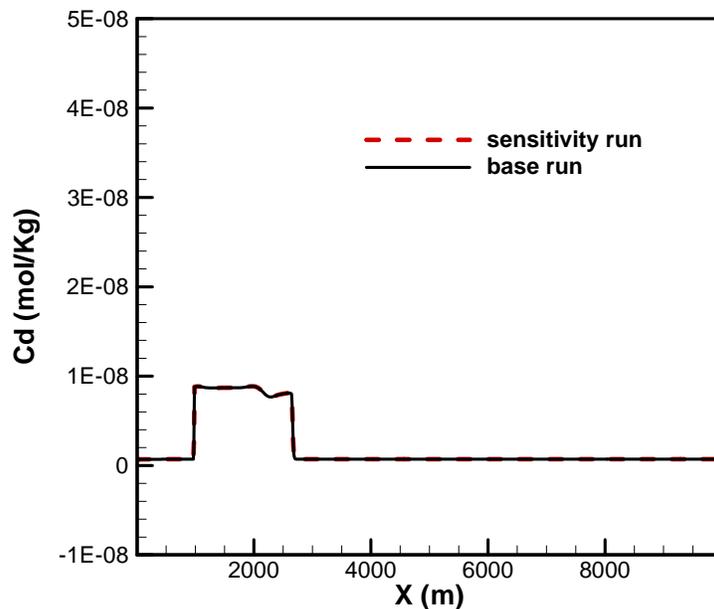


Figure 10. Spatial distribution of dissolved Cd concentration at 200 years in the base-case run and a sensitivity run that inherits all the chemical reactions in the base-case run but with Cd concentration in the leaking brine at 10^{-6} mol/kg, about 1.5 order of magnitude higher than that in the base run.

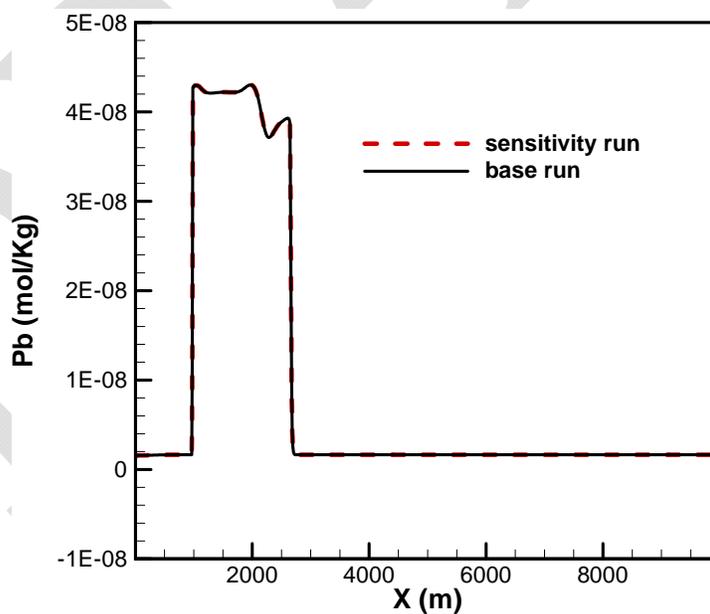


Figure 11. Spatial distribution of dissolved Pb concentration at 200 years in the base-case run and a sensitivity run that inherits all the chemical reactions in the base-case run but with Pb concentration in the leaking brine at 10^{-5} mol/kg, about 1.75 order of magnitude higher than that in the base run.

2.2 GLOBAL SENSITIVITY ANALYSIS

We performed a relatively simple assessment of the sensitivity of some input parameters, as described in the previous section. Here, for a better understanding of the controlling processes, we discuss a more rigorous global sensitivity analysis of the parameters associated with the chemical processes. The performance measures for the global sensitivity analysis are as follows, for 10 simulated time periods (20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 years):

- Volume of aquifer reaching pH < 6.5
- Volume of aquifer reaching TDS > 500 mg/L
- Volume of aquifer reaching concentrations of As > MCL (established by the U.S. Environmental Protection Agency [EPA], 1.33e-7 mol/kg)
- Volume of aquifer reaching concentrations of Cd > MCL (4.05e-8 mol/kg)
- Volume of aquifer reaching concentrations of Pb > MCL (7.24e-8 mol/kg)

The global sensitivity analysis of the 1-D model was performed using PSUADE (Tong, 2010) in combination with TOUGHREACT version 2 (Xu et al., 2012). PSUADE (Problem Solving environment for Uncertainty Analysis and Design Exploration) is a software package for various uncertainty quantification (UQ) activities, such as uncertainty assessment (UA), sensitivity analysis (SA), parameter study, and numerical optimization.

2.2.1 Sampling

By design, the 1-D model is rather simple in terms of characterization of flow and transport processes, and focuses largely on the geochemical reactions, for the purpose of conducting global sensitivity analysis only on geochemical parameters. Table 3 lists the considered input variables for global sensitivity analysis and their ranges. Note that the sorption factor is a generalized parameter for characterizing the sorption capacity of the aquifer relative to the base case for the High Plain Aquifer model, which considers (as sorbing minerals) 2.85 vol.% illite, 1.36 vol.% smectite, 1.7 vol.% kaolinite, 0.6 vol.% iron hydroxide in the aquifer sediments. A Latin hypercube sampling method was adopted, and a total of 1000 samples were generated corresponding to 1000 simulations of the 1-D numerical model.

2.2.2 Sensitivity analyses

In this section, we discuss the sensitivity of performance measures to those parameters listed in Table 3. Note that if the concentrations of trace metal do not exceed their perspective MCL, the volume of aquifer with concentration > MCL will be zero. Simulation results showed that this was the case for Pb and Cd in most simulations. In 1,000 simulations, we obtained only 4 non-zero volumes for Cd and 81 non-zeros for Pb. Such small numbers of non-zero values prevent us from performing meaningful UQ for Pb and Cd. Therefore, we limit our analyses to the volume of aquifer with pH < 6.5, the volume of aquifer with TDS > 500 mg/L, and the volume of aquifer with concentration of As > MCL.

Table 3: Input variables and ranges for global sensitivity analysis

Number	Name	Description	Minimum	Maximum	Distribution
1	Pbbrine	Pb concentration in the leaking brine	-8.5	-5	Log- uniform
2	Cdbrine	Cd concentration in the leaking brine	-9.0	-6.0	Log-uniform
3	Asbrine	As concentration in the leaking brine	-9.0	-5.0	Log-uniform
4	Clbrine	Cl concentration in the leaking brine	-2	1	Log-uniform
5	calcitevol	Calcite volume fraction	0	0.2	Log-uniform
6	sorptionfactor	The variation in the sorption capacity of the aquifer	-2	2	Log-uniform

Figure 12 shows the global sensitivity indexes for the volume of aquifer with pH<6.5, indicating that the sorption factor is the most sensitive parameter. It is not surprising that the volume of aquifer with pH<6.5 is not sensitive to the concentrations of As and Cl in the leaking brine. It should be noticed that the calcite volume might have an effect on the resultant lowest pH, but it appears that it does not affect the volume of aquifer with pH<6.5. The volume of aquifer with TDS > 500 mg/L is also most sensitive to the sorption factor (Figure 13), because this parameter greatly affects the area with low pH where dissolved CO₂ is elevated and the dissolution of minerals occurs from the depressed pH. This process can easily cause the TDS to increase above 500 mg/L because the initial water already has a TDS of 430 mg/L. The volume of aquifer with TDS > 500 mg/L is also sensitive to the concentration of Cl in the leaking brine. Figure 14 shows that the volume of aquifer with concentration of As > MCL is exclusively sensitive to the sorption factor, because the As concentration in the aquifer is controlled by adsorption/desorption via surface complexation. It is not sensitive to the As concentration in leaking brine, which is consistent with the results shown in Figure 9. As discussed earlier, the effect of the concentration of As in leaking brine on the resulting As concentration in the aquifer is largely overshadowed by adsorption/desorption.

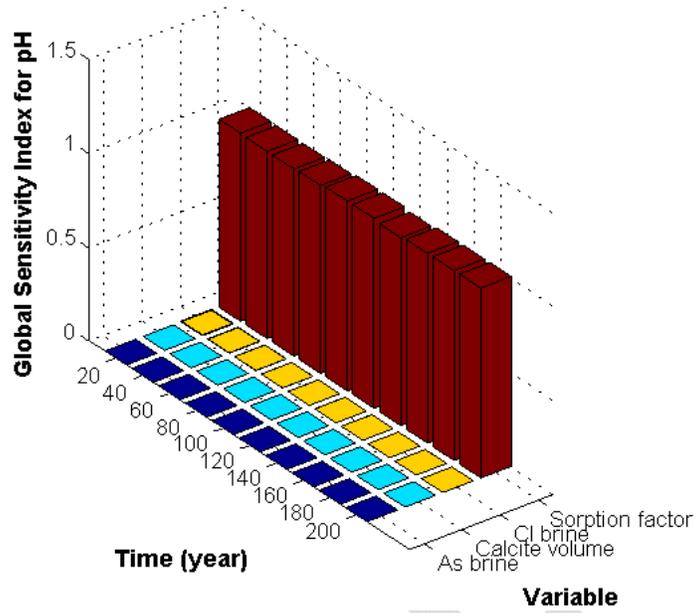


Figure 12. The Sobol global sensitivity index for the volume of aquifer with pH<6.5.

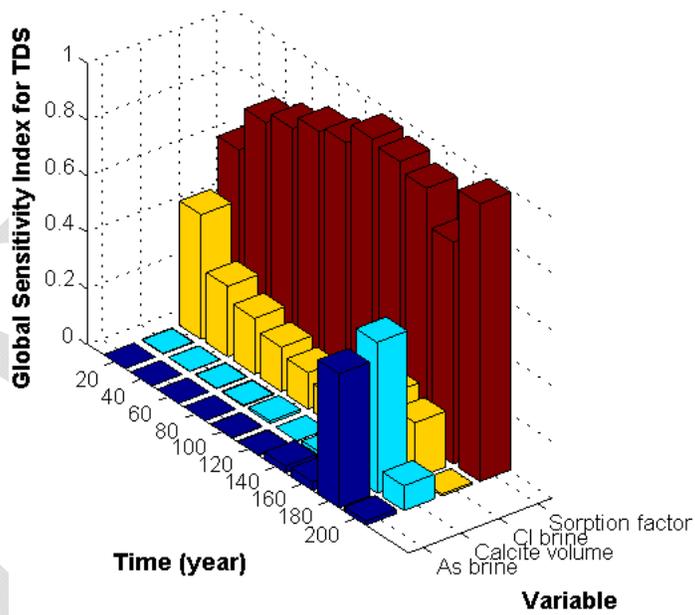


Figure 13. The Sobol global sensitivity index for the volume of aquifer with TDS>500 mg/L.

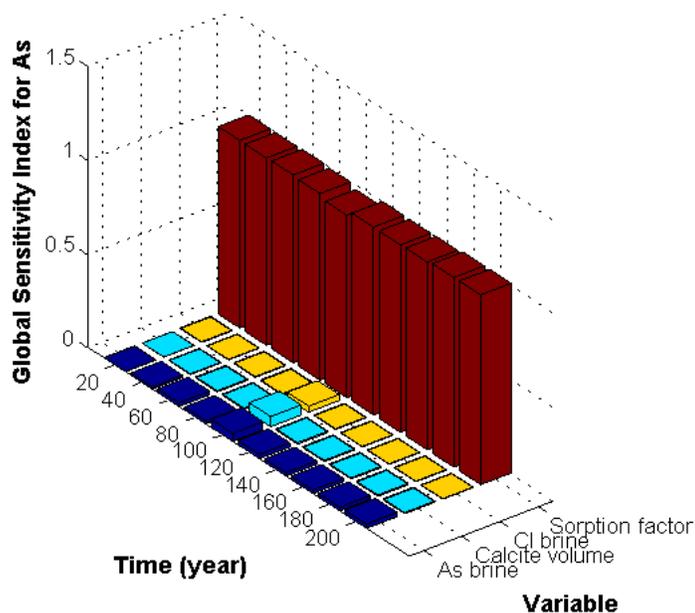


Figure 14. The Sobol global sensitivity index for the volume of aquifer with concentration of As > MCL.

2.3 ROM DEVELOPMENT

As mentioned in the previous section, out of 1,000 simulations, we obtained only 4 non-zero groundwater volumes affected by Cd and 81 non-zero results for Pb. This limited number of significant results does not allow us to derive a ROM for the volume of aquifer with Pb and Cd concentration higher than their perspective MCLs. We therefore developed ROMs to estimate the volume of aquifer with pH < 6.5, the volume of aquifer with TDS > 500 mg/L, and the volume of aquifer with concentration of As > MCL.

The input parameters for the ROM are variable #3, #4, #5, #6 in Table 3. For each output variable (i.e., volume of groundwater affected by pH, TDS, and As), the ROM is composed of 20 quadratic polynomial functions, each one representing a time spot from 10 to 200 years with equal 10-year increments. Those polynomial functions are derived by regression of the volumes calculated in the 1,000 simulations. Figure 15 shows the R² for each ROM. In general, the estimated R² values are acceptable, as most of them are higher than 0.8. However, values significantly below 0.8 were found for the ROM of the volume of aquifer with TDS > 500 mg/L for times close to 200 years (Figure 15). The main reason for this finding is that 500 mg/L is such a low threshold value that different combinations of input parameters often lead to the same predicted impacted groundwater volume, which decreases the accuracy of the regression analysis on which the derivation of the ROM is based. Figures 16 and 17 show the comparison between emulated volumes of aquifer with TDS > 500 mg/L with ROM and simulated ones with numerical models for 70 and 200 years, respectively. Note that for 200 years, the range of volume is very small, and simulated results have only three different values, which make emulation very difficult. Figure 16 and 17 show the comparison of emulated and simulated volumes of aquifer with pH < 6.6 and As concentration > MCL respectively. Emulated results match closely with simulated results, indicating that the derived ROM is reliable.

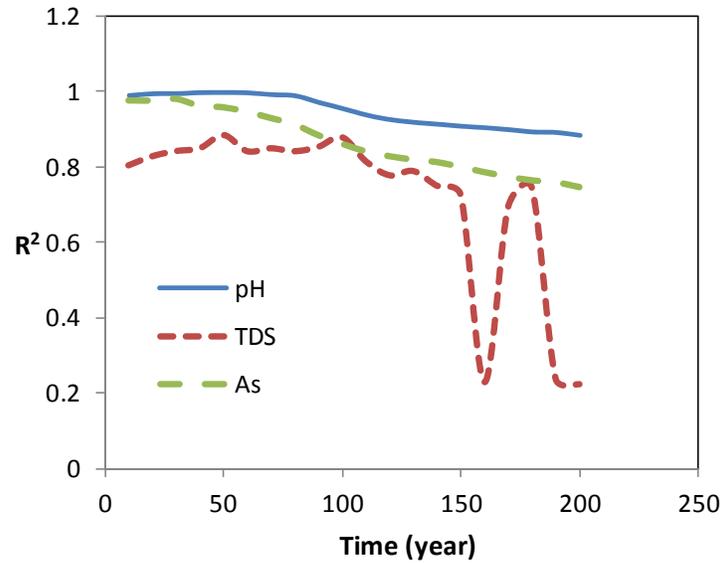


Figure 15. The R² of ROMs for the volume for aquifer with pH<6.5, TDS>500 mg/kg, and As concentration >MCL.

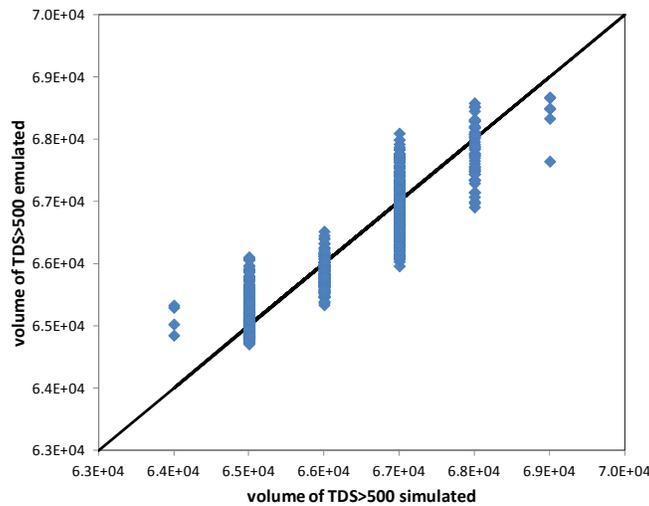


Figure 16. Comparison of emulated volumes of aquifer with TDS>500 mg/L with ROM and simulated ones with numerical models for the 70 years. The solid line represents a one-to-one match between emulated and simulated.

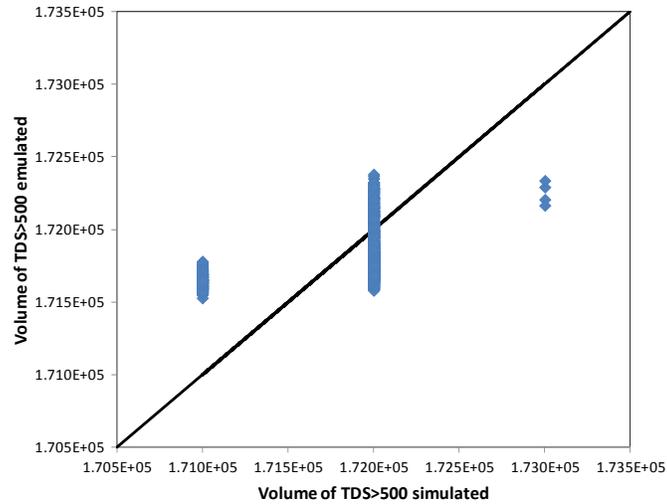


Figure 17. Comparison of emulated volumes of aquifer with TDS > 500 mg/L with ROM and simulated ones with numerical models for the 200 years. The solid line represents a one-to-one match between emulated and simulated.

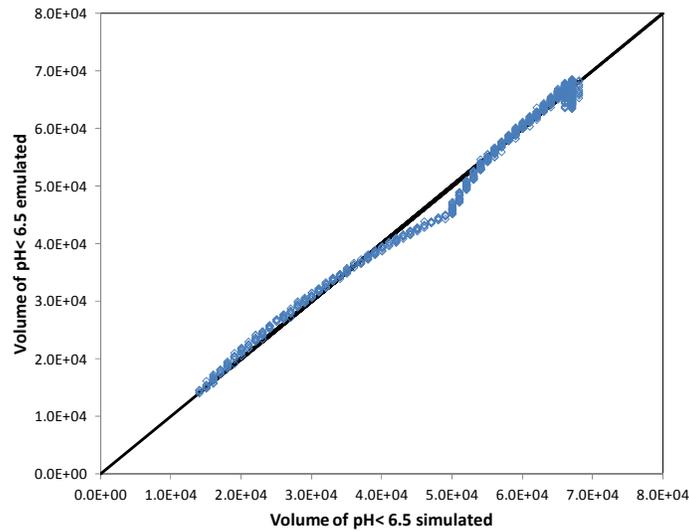


Figure 18. Comparison of emulated volumes of aquifer with pH < 6.5 with ROM and simulated ones with numerical models for the 70 years. The solid line represents a one-to-one match between emulated and simulated.

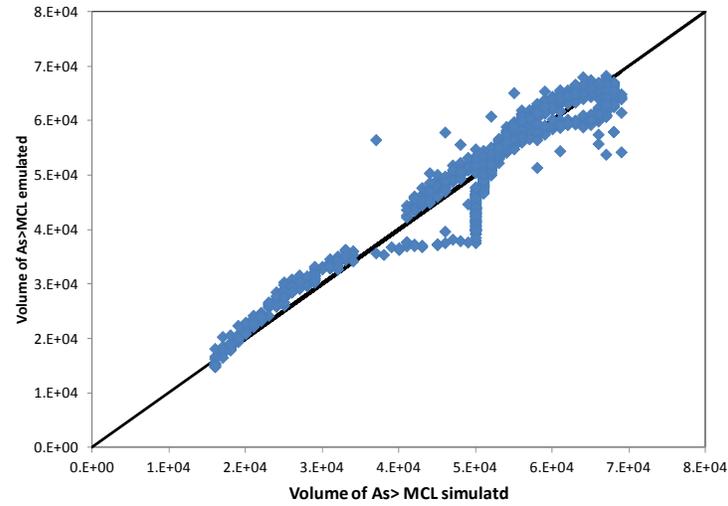


Figure 19. Comparison of emulated volumes of aquifer with As concentration >MCL with ROM and simulated ones with numerical models for the 70 years. The solid line represents a one-to-one match between emulated and simulated.

DRAFT

3. LINKING FUNCTION

3.1 INTRODUCTION AND BACKGROUND

Computer simulations of physical systems are widely used for natural resource management and for supporting the design and implementation of environmental policies and decisions. These simulations are essential when laboratory or field experiments are too time consuming, expensive, or practically impossible to conduct. This last case is not uncommon, especially when the spatial and/or the time scales of the natural system are large. Computer simulations rely upon a mathematical model, a set of mathematical expressions describing the behavior of the physical system, implemented in a computer code. Despite the constant advance in computational resources, modern computer simulations applied to solve problems in the natural sciences and engineering disciplines still require a great computational effort, in part because progress in computational power goes hand-in-hand with the ability to model higher degrees of complexity. Complexity is in fact necessary to take into account various physical and chemical processes, often coupled, in large heterogeneous environmental systems. Computational power is particularly demanding for numerical experiments requiring multiple runs of the same computer simulation with various combinations of input parameters over a range of interest. Such numerical experiments are essential for model calibration, optimization, sensitivity, and uncertainty analysis, as well as for risk assessment.

Since the introduction of computer modeling, there has been a constant discussion about the dilemma between complexity and simplicity. In the field of groundwater modeling, for example, this is still a debated topic (Haitjema, 2006; Hill, 2006; Gómez-Hernández, 2006; Clement, 2011; Doherty, 2011). Arguments in favor of simple models include computational efficiency, numerical stability, and the fact that their outputs are easily understandable and interpretable (e.g., Lee, 1973; Grimm et al., 1999). However, oversimplification caused by neglecting important aspects, such as spatial heterogeneity or the effects of relevant physical and chemical processes, may result in only a partial or inaccurate representation of the real system. On the other hand, complex models overcome this limitation, allowing us to simulate complex processes as well as to include spatial heterogeneity and a larger number of input parameters. Stochastic variability is needed for estimating the uncertainty associated with model predictions and their sensitivity to the different input parameters. The counterpart of improving the realism of the simulation is the computational effort, which, in some cases, can become prohibitively large. Moreover complex models may be numerically unstable, and model calibration, which requires the adjustment of the input parameters to match experimental observations, may be very difficult due to the higher number of input parameters. A more comprehensive discussion on issues related to simple versus complex models can be found in Van Nes and Scheffer (2005).

In recent years, a new simulation approach has emerged, which combines both complex and simple models in the same analysis, so as to benefit from the advantages of both. Van Nes and Scheffer (2005), for example, proposed a method for understanding the role of different parameters of an ecological model based on the comparison of the outputs of complex and simplified models. Combining complex and simple models is also effective for reducing the computational time of parameter estimation problems in calibrating reservoir simulation models to dynamic data (e.g., Aanonsen and Eydinov, 2006; Lødøen and Tjeelmeland, 2010; Scheidt et al., 2011). With this approach, known as the multiscale parameter-estimation method, a series of models with different resolutions are developed through upscaling of an initial fine-scale,

complex model. Significant reduction of computational effort and increase in accuracy is achieved when the coarse models are initially used for estimating large-scale corrections of the input parameters, and then their calibrated values are adopted as initial estimates for final adjustment at the fine scale.

Pairing of complex and simple models is also widely used in the surrogate modeling practice, which refers to the development and use of simplified and fast models, also known as ROMs, in lieu of the original simulation models. The two broad categories of ROMs include response surface models, which empirically approximate the original model response, and low-fidelity models, which are physically based (Razavi et al., 2012). The ROM described in the previous sections of this report is an example of the first category. Low-fidelity models represent a simplified but faster-to-run alternative to a more detailed complex model, which is usually designated as a “high-fidelity” model, since it is assumed to provide the most faithful representation of the real system. Different approaches can be taken for the simplification of a high-fidelity model—such as the use of a coarser spatial or temporal discretization, parameter upscaling, a lower dimensional representation (i.e., two-dimensional instead of three-dimensional), the introduction lumped parameters, and the exclusion of certain physical or chemical processes. High-fidelity models can be efficiently used in conjunction with lower-fidelity models to increase the computational efficiency of optimization problems. In this approach, known as “multi-fidelity” or “variable-fidelity” optimization (e.g., Gano et al., 2006; Forrester et al., 2007; Sun et al., 2010; Berci et al., 2011), the discrepancies between high-fidelity and low-fidelity models are most commonly addressed by correction functions. They are usually linear functions or second-order polynomials (e.g., Madsen and Langthjem, 2001; Viana et al., 2009; Sun et al., 2010), but more complicated and flexible relationships, such as kriging and neural networks, have been applied (Gano et al., 2006; Forrester et al., 2007).

In this report, we propose a methodology to couple simple and complex models to approximate the response of a complex model using the responses of two or more simple models. By exploiting the advantages of both complex and simple models, this methodology significantly reduces the computational time without compromising the realism and the accuracy of the simulations. Particularly relevant for the development of the proposed approach are the work of Kennedy and O’Hagan (2000) and Doherty and Christensen (2011). Kennedy and O’Hagan used Bayesian analysis and proposed a first-order autoregressive model to link the output of a complex code to the correspondent output of a simplified version. On the other hand, the recently proposed method by Doherty and Christensen (2011) used the subspace concept to develop a paired complex/simple model approach, in which the predictive bias of the simple model is corrected through a function, assumed to be linear for simplicity, which describes the best fit of the simple model outputs to those of the complex model.

The proposed methodology can be viewed as an extension of these two approaches; the similarity lies in the fact that, as with the previous authors, we propose a mathematical function, hereafter referred to as the *linking function* (LF), to “link” the outputs of models with different levels of complexity. However, our methodology is more flexible, since it does not require any particular shape or dimensionality of the linking function, so that it can handle complex and nonlinear relationships. For this reason, neither simple nor complex models are required to have the degree of smoothness required in the method proposed by Kennedy and O’Hagan (2000). In short, our methodology is more general, can be applied to different simplification strategies, and allows the use of multiple simple models, each of which are able to focus on a specific aspect of

the real system. Moreover, these simple models do not need to be physically based, but rather can be a response surface surrogate. As described in Section 6.4, the linking functions developed to combine the responses of two ROMs that were developed to investigate the impact of CO₂ and brine leakage from a single wellbore on groundwater quality in the High Plain Aquifer. Each of these two ROMs represents a simplified representation of a high-fidelity model. The LBNL ROM considers all relevant chemical parameters, but it is rather simple in terms of flow and transport behavior, whereas the LLNL ROM considers parameters related to heterogeneous multiphase flow and transport, but neglects important geochemical processes. Applying the linking function approach allows us to approximate the output of the high-fidelity model by combining the outputs from these two ROMs.

3.2 THE LINKING FUNCTION METHOD

The following description is the methodology used for development of the linking function. Let Z_c be a complex model considering all the known processes acting in a real-world system and the explanatory parameters. For a specific set of input parameters \mathbf{x} , its output Y_c can be calculated by running the complex model such that $Y_c = Z_c(\mathbf{x})$. Also, let $\mathbf{y}_s = [y_{s1}, y_{s2}, \dots, y_{sN}]$ be a vector of outputs from a number N of simple models z_{si} , each of which represent a lower-fidelity approximation of the same system. If we assume that (1) each of the simple models shares some basic features with the complex model, meaning that the input parameters for the simple models are a subset of those of the complex model, and that (2) the outputs of all modes are scalar and not multivariate time series responses, similar to Kennedy and O'Hagan (2000), the output from the complex model Z_c can be defined as follows:

$$Y_c = g(\mathbf{y}_s, \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \quad i = 1, 2, \dots, N \quad (1)$$

where g is a mathematical function, hereafter referred to as the linking function, $\boldsymbol{\beta}$ is a vector of unknown parameters of the linking function, and $\boldsymbol{\varepsilon}$ is a random errors vector. As previously stated, the function $g(\mathbf{y}_s, \boldsymbol{\beta})$ can take various mathematical forms, which must be defined on a case-by-case basis.

The linking function methodology can be implemented through the following steps:

- 1) Develop a complex model Z_c for the considered system
- 2) Develop the simple models z_{si}
- 3) Design a numerical experiment to perform multiple runs of the complex model with different sets of input parameters and calculate a vector of outputs \mathbf{Y}_c
- 4) Perform multiple runs of the models z_{si} and calculate, for each run, a vector \mathbf{y}_s . To ensure the correspondence between the outputs from the complex and simple models, the input parameters of the simple models must be a subset of those of the complex model. For instance, if the j -th run for the complex model of the numerical experiment was performed with a set of n -input parameters x_i ($i = 1, n$), then the set of m -input parameters ($m < n$) for the j -th run of simple model z_{s1} is x_i ($i = 1, m < n$). If another simple model z_{s2} is considered, which takes into account the processes and parameters omitted in z_{s1} , the set of input parameters for the j -th run of z_{s2} is x_i ($i = m+1, n$).

- 5) Identify the mathematical function that represents the best fit between outputs from the simple and the complex outputs. This mathematical function is the *linking function*, and it is the core to this proposed methodology. This step consists of a regression analysis in which the dependent variable is the output from the complex model Y_c , and the independent variables are corresponding outputs y_{si} of the simple models. The shape and the coefficients of the linking function can be estimated with the least-square-regression method or other fitting procedures (e.g., kriging).
- 6) Run the simple models with a specific set of input parameters for the problem of interest, and use the linking function to emulate the response of the complex model

This methodology can significantly decrease the computational time as well as provide accurate results. By addressing the discrepancies between simple and complex models, the linking function is in fact an effective way to retain the level of realism and detailed information associated with complex models, while at the same time avoiding the long computational times usually associated with running such models. Although the implementation of the methodology requires some computational cost, the development and application of the linking function can be still more advantageous for some types of numerical analysis, such as uncertainty quantification of the complex model output and sensitivity analysis of its several input parameters. The traditional way to perform these analyses requires many runs of the complex model, which can be time consuming and computationally intensive. Instead, since the number of simple models is expected to be much smaller than the number of parameters in the complex model, a relatively small number of complex model runs are required to develop the linking function. Once the linking function is developed, uncertainty quantification and sensitivity analysis can be performed by running only the faster simple models.

Another fundamental advantage of this methodology is that the simple models do not need to be physically based. They can be surrogate models, such as a previously developed response surface models. With more computational effort, polynomial response surface surrogates of the physically-based simple models can even be developed at the same time as the linking function. This will translate to even faster prediction of the complex model response with the linking function. However, note that when surrogate models are used as input in the linking function, a further source of error is introduced, related to the accuracy of the surrogate models in emulating the outputs of simple models.

3.3 DEVELOPMENT OF LINKING FUNCTION

In NRAP, to develop a ROM for estimating the impact of CO₂ and brine leakage in shallow aquifers that can account for all the relevant parameters, we needed to develop a very complex model that considers all flow, transport, and chemical processes. However, such a model would be so computationally expensive that it would be practically impossible to generate a ROM based on a number of simulations of it. Therefore two ROMs were developed on the basis of the hydrogeological and geochemical conditions of the High Plain Aquifer. The first ROM was derived from a model that assumed heterogeneous flow and transport with a very complex leakage scenario while the second ROM was based on models that were very simple in terms of characterization of flow and transport, but very complex in terms of taking into account realistic chemical reactions in the aquifer (Section 3.1).

In this section, we will discuss the development of a linking function intended to link these two ROMs. First, we developed two simple models: a 1-D model with homogenous flow and transport and a single leakage point, but including as many chemical reactions as possible; and a 2-D model considering aquifer heterogeneity but no reactions. Second, a complex model that incorporates all the parameters and physical and chemical processes of both simple models was developed. Third, according to the procedure illustrated in the previous section, multiple runs of simple and complex models were conducted, and a linking function was finally developed based on simulation results.

3.3.1 Simple 1-D model

The simple model used for developing the linking function is exactly the same as the model described in Section 3.1. The input parameters are also the same as those described in Section 3.2.

3.3.2 Simple 2-D model

Hydrogeological setting

The cross-sectional 2-D model targets the potential impact of CO₂ injected into an aquifer with lithological, hydrogeological, and geochemical properties based on the High Plain Aquifer. Specifically, the lithological characterization is based on the lithological descriptions of 48 wells located in Haskell County, in southwest Kansas. The source of these data is the Water Well Completion Records (WWC5) Database (Kansas Geological Survey, 2012). These data are the same as those on which the LLNL 3-D model was developed, providing the link between our model and that of LLNL. The lithologies include different types of unconsolidated sediments with a highly heterogeneous granulometric distribution, typical of a fluvial depositional environment. For simplicity, the original lithological descriptions were classified into two hydrostratigraphic units on the basis of grain size (coarse/fine) and permeability (high/low) as reported in the scientific literature (e.g., Bear, 1972). The lithologies included in each of these units are provided in Table 4.

Table 4. Lithologies associated with the hydrostratigraphic units of the synthetic aquifer

Hydrostratigraphic Unit	Lithologies as in the WWC5	Mean Length (horizontal direction)	Mean Length (vertical direction)	Volumetric Proportion
Unit 1	Sand, coarse sand, medium sand, sand with gravel, gravel with sand, medium gravel, gravel, coarse gravel	717.7 m	8.3 m	0.60
Unit 2	Fine sand, very fine sand, silty sand, silt, silty clay, shale, sandstone, caliche, gypsum rock, clay, limestone	478.5 m	5.6 m	0.40

The distribution of these two units was simulated with the transition probability method (T-PROGS; Carle, 1999), which was also used for the development of the LLNL 3-D model. Heterogeneous fields generated with T-PROGS are based on the transition probabilities between different categories and on a single Markov Chain equation in each direction. Transition

probabilities are defined as the probability that a certain category j occurs at the location $\mathbf{u}+\mathbf{h}$ conditioned to the occurrence of another category i at the location \mathbf{u} . Here \mathbf{u} and \mathbf{h} are a location and a movement vector, respectively. One advantage of this methodology is the increased realism of the simulations, making it thereby possible to account for observable geological features such as mean lengths and juxtapositional tendencies. A thorough description of the algorithm can be found in Carle and Fogg (1996, 1997), Carle et al. (1998), and Carle (1999). T-PROGS simulations of the aquifer heterogeneity were conducted with mean lengths and volumetric proportions for the two different hydrostratigraphic units estimated from the analysis of their spatial distributions in the 48 wells. Values for these two parameters are given in Table 4.

The aquifer is assumed to be confined, and the mean groundwater flow is from east to west with a hydraulic gradient of 0.003. Aquifer thickness is uniform and equal to 240 m, which corresponds to the average thickness of the High Plain Aquifer in Heskell County. In this model, chemical species are treated as conservative species, and their initial concentrations in the aquifer and in the leaking brine are listed in Tables 1 and 2.

The simulated domain is rectangular, representing a cross section of the aquifer (oriented in the E-W direction) of length equal to 10,000 m and thickness equal to 240 m. The spatial distribution of the two hydrostratigraphic units (Table 4) corresponds to one unconditional realization of the T-PROGS geostatistical model. The interpolation grid is composed of rectangular cells with constant dimensions equal to 100 m and 5 m in the x-direction and z-direction, respectively (Figure 20a). In the implementation of the multiphase and transport models described in the next sections, different hydrological parameters (i.e., permeability, porosity, etc.) were assigned to each unit.

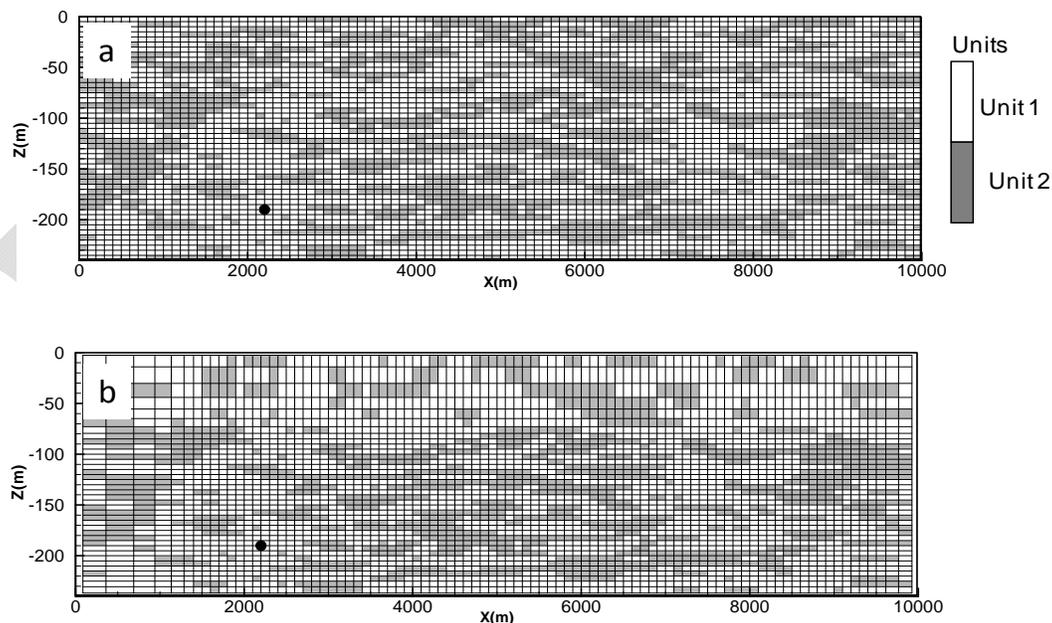


Figure 20. Heterogeneous distribution of two hydrostratigraphic units generated with T-PROGS (a). Numerical model mesh used for the TOUGHREACT simulations (b). The black circle indicates the location of the CO₂ and brine leakage point.

Model implementation and base-case results

Multiphase flow and reactive transport were simulated with the TOUGHREACT 2.0 code (Xu et al., 2012). The numerical mesh and the T-PROGS interpolation grid are identical in more than 70% of the simulated domain, with some differences in proximity of the boundaries, where the numerical mesh is coarser (Figure 20b). The coarsening of the mesh, to save computational time, was done by multiplying element dimensions by a factor of 1.2, starting from the coordinates $x = 1500$ m toward the left boundary, and from $z = -80$ m toward the top of the domain. The resulting mesh is composed of a total of 3,549 elements. Before choosing this mesh design, trial runs were also performed with a regular mesh composed of elements with the same dimensions as in the interpolation grid, to ensure that coarsening the mesh at the boundaries does not affect simulation results in the area where leakage occurs and where the CO₂ and brine plume move.

Initial pressure and boundary conditions for flow are shown in Figure 21. We imposed Dirichlet conditions (constant hydraulic head) at the left and right boundaries, while no-flow boundary conditions were applied at the top and bottom of the domain. A preliminary gravity equilibration run, without CO₂ and brine injection, was run long enough to reach quasi-steady-state conditions (Figure 21). The resulting flow field was used as the initial flow field for the CO₂ and brine leakage simulations. These runs were conducted at constant temperature 17°C.

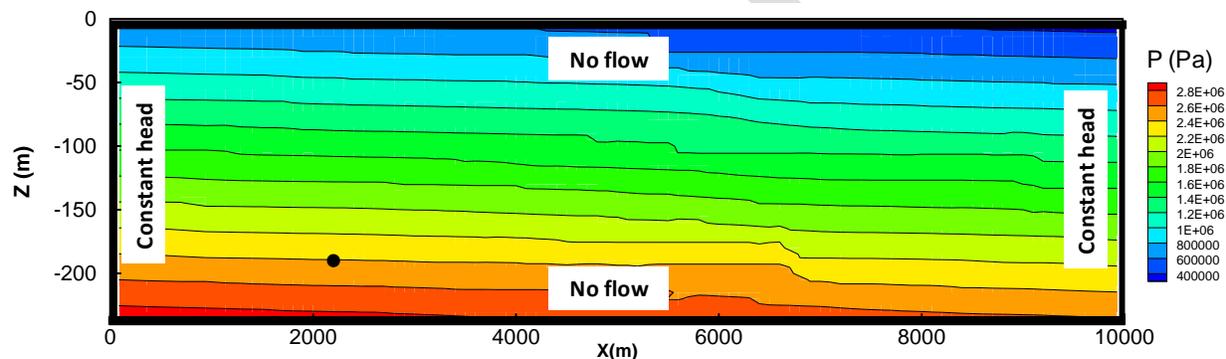


Figure 21. Initial and boundary conditions for the simple 2-D model. The same applies for the complex model.

Leakage of CO₂ and brine from the wellbore was simulated by assuming a point source at the point of coordinates (2200 m, -190 m) and a duration of 200 years with variable leakage rates. The maximum values of these leakage rates are plotted as a function of simulation time in Figure 22. These maximum values were then multiplied by a factor during the development of the linking function. The CO₂ rates sharply increase during the initial 5 years and then oscillate, with variations ranging from 0.039 kg/s to 0.046 kg/s. The brine leakage rates are more stable, with very little variation around an average of 0.012 kg/s. Leakage rates are derived from the first-generation ROM developed for wellbore leakage.

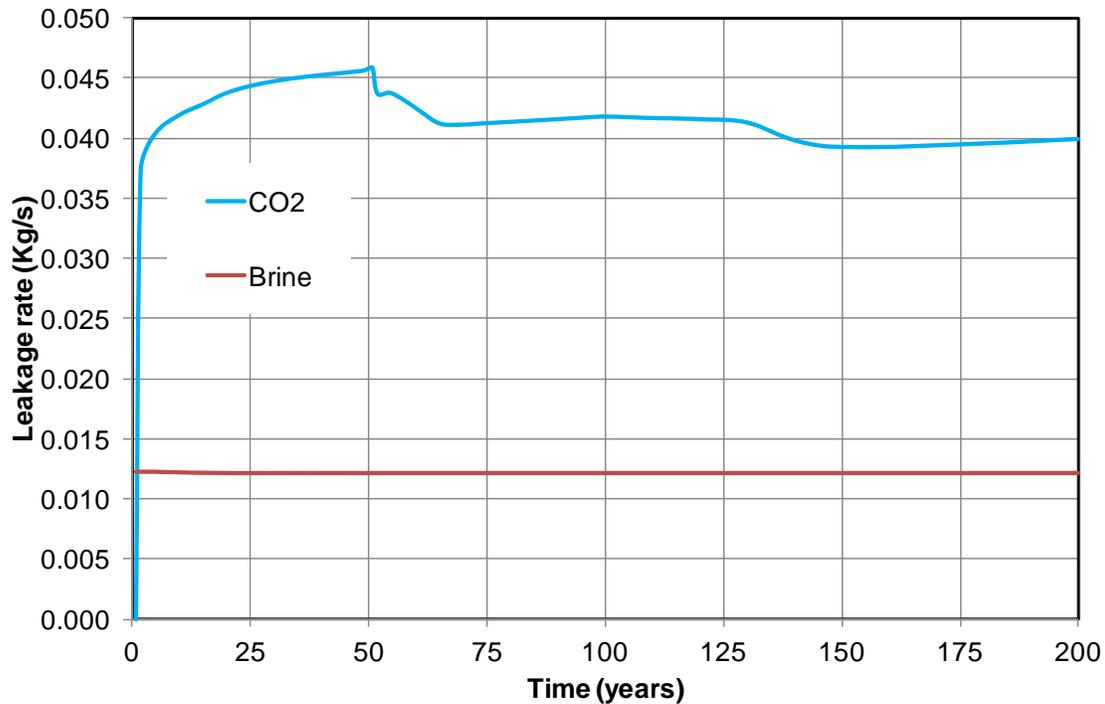


Figure 22. Maximum CO₂ and brine leakage rates over time.

Input parameters for the base case are reported in Table 5. The results for a base-case run of the simple 2-D model are shown in Figure 23. The base-case simulation results were used to understand the system behavior and to manually test the sensitivity of the model to the different input parameters. After 200 years of continuous release of brine and CO₂ from the leakage point, the area with lowered pH values and the plumes of three considered metals (As, Pb, and Cd) moved about 7 km downgradient. As expected, the major role in determining the shape of these plumes is played by the heterogeneous distribution of the two hydrostratigraphic units, with the plume following preferential flow paths to the location of the highest permeable unit.

Table 5. Input parameters for the simple 2-D model base case

Parameter	Base Case Value
Porosity (unit 1)	0.250
Porosity (unit 2)	0.330
Rock density (unit 1)	2400 kg/m ³
Rock density (unit 2)	2400 kg/m ³
Permeability (unit 1)	3.162x10 ⁻¹¹ m ²
Permeability (unit 2)	3.162x10 ⁻¹⁷ m ²
Van Genuchten parameter m (unit 1)	0.655
Van Genuchten parameter m (unit 2)	0.190
Van Genuchten parameter alpha (unit 1)	5.62x10 ⁻⁵ m ⁻¹
Van Genuchten parameter alpha (unit 2)	1.51x10 ⁻⁵ m ⁻¹

3.3.3 Complex model

The complex model for simulating the impact of CO₂ and brine leakage on the aquifer consists of a combination of the two simple models described in Sections 4.3.1 and 4.3.2. While inheriting the model setup, hydrogeological parameterization, and leakage function of CO₂ and brine in the simple 2-D model, the complex model also incorporates all the chemical reactions that were used in the simple 1-D models. This model is expected to provide the most accurate representation of the natural system, taking into account uncertainties in flow, transport, and chemical processes. Figure 24 shows the plumes of pH, As, Pb, and Cd for the base-case simulation. Input parameters are reported in Tables 2 and 5.

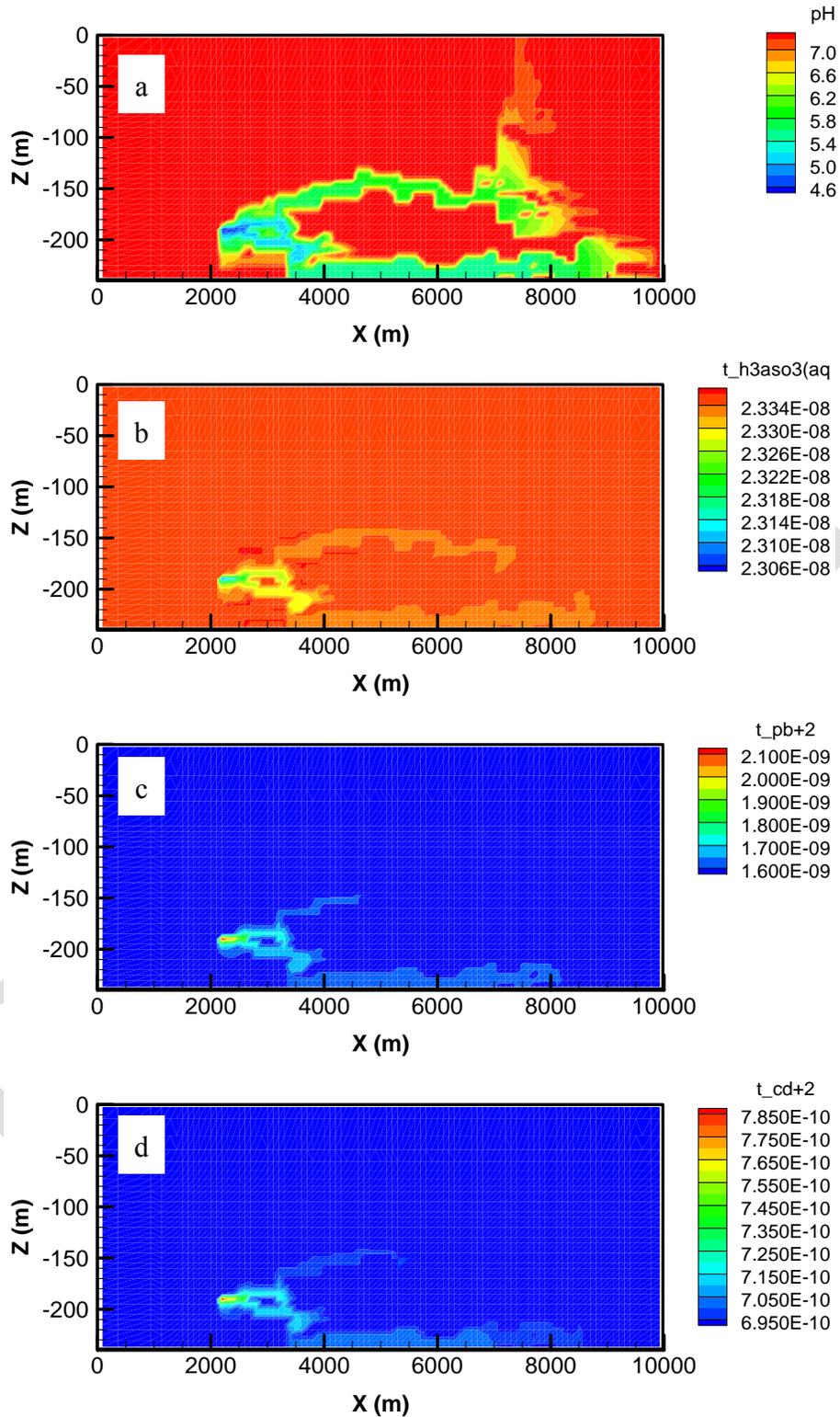


Figure 23. Simple 2-D model base-case simulation results after 200 years of continuous leakage (unreactive transport). pH distribution (a), AsO₃ concentration (b), Pb²⁺ concentration (c), Cd²⁺ concentration (d). The leakage point is at X=2200 and Z=-190 m.

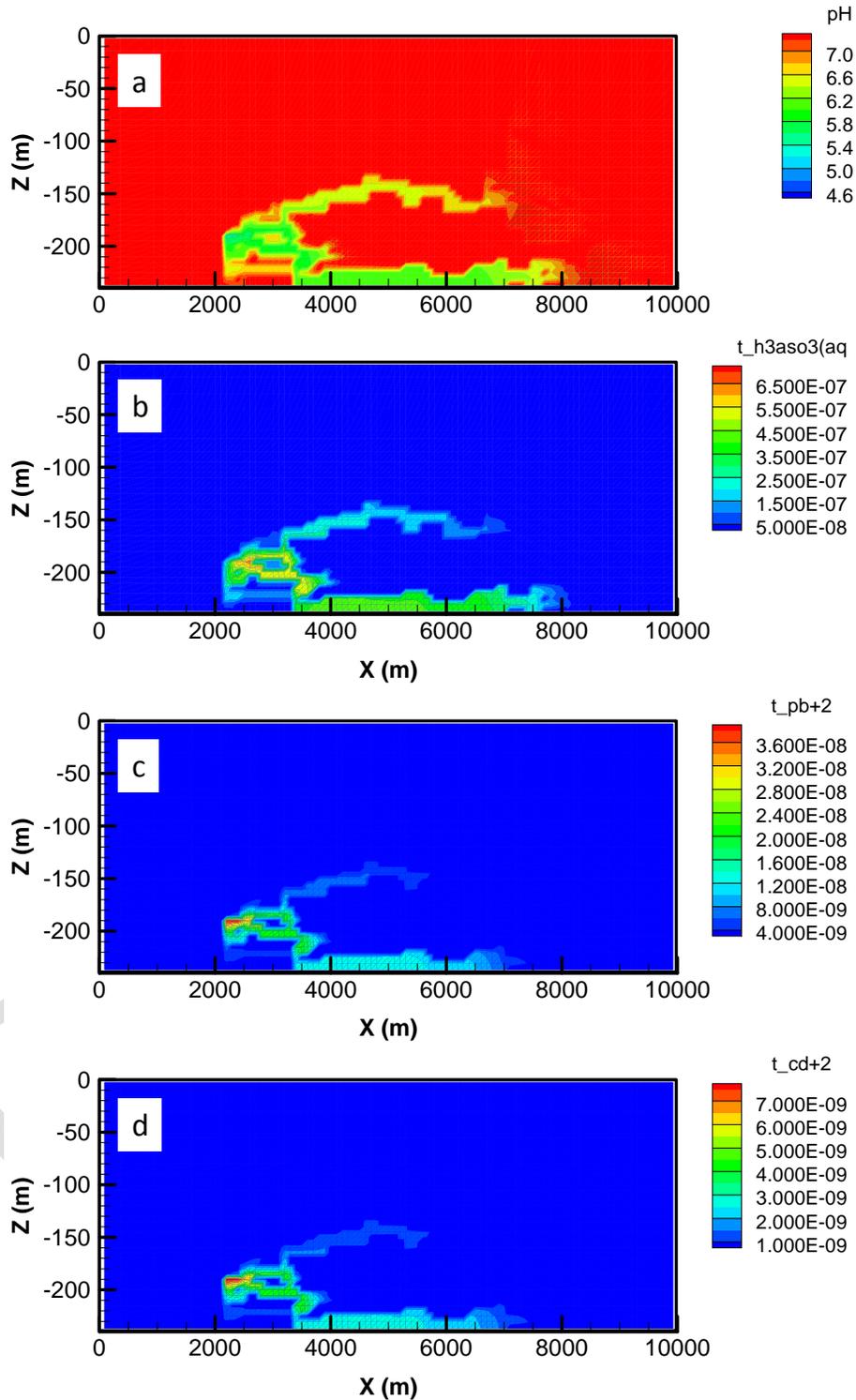


Figure 24. Complex 2-D model base-case simulation results after 200 years of continuous leakage (reactive transport). pH distribution (a), AsO₃ concentration (b), Pb²⁺ concentration (c), Cd²⁺ concentration (d). The leakage point is at X=2200 and Z=-190 m.

3.4 LINKING FUNCTIONS AND RESULTS

To develop the linking functions between the simple and complex models, we considered the following output variables for 20 simulated time periods (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, and 200 years):

- Volume of aquifer reaching pH < 6.5
- Volume of aquifer reaching TDS > 500 mg/L
- Volume of aquifer reaching As concentrations > MCL (established by the EPA, 1.33e-7 mol/kg)
- Volume of aquifer reaching Cd concentrations > MCL (4.05e-8 mol/kg)
- Volume of aquifer reaching Pb concentrations > MCL (7.24e-8 mol/kg)

Details of the work conducted for each of the steps described in Section 4.2 are summarized below. The entire procedure was implemented in a GNU Octave v.3.6.1 script that was written for this purpose.

- Steps 1 and 2. Development of the complex and simple models as described in Section 4.3.
- Steps 3 and 4. Design of a numerical experiment and perform multiple runs of the simple and complex models. The numerical experiment involving multiple runs of the simple and complex models with different sets of input parameters was designed by using a quasi-random sequence (LP τ , Shukhman, 1994) to generate 450 sample points in the input parameter space. The statistical distributions of the input parameters are assumed to be uniform. Details on parameter ranges are given in Table 6. Parameter distributions correspond to the distributions used for the generation of the LNLN and LBNL ROMs.
- Step 5. Estimate the linking functions for each output variable and for 20 simulation times. The linking functions are estimated through a least-squares regression analysis in which the outputs from the simple 1-D and 2-D models are used as independent variables in order to fit the outputs from the complex model. For all the output variables and for all the simulation times, a second order polynomial was found to provide a close match between the simple and complex model outputs. An example of the shape of this polynomial function is shown in Figure 25. The coefficients of all the generated linking functions are reported in Appendix A. The goodness of fit for the developed linking functions is analyzed by calculating the coefficient of determination (R^2). Taking into account all the linking functions, the calculated R^2 values are between 0.701 and 0.998, with the majority of values higher than 0.900. In general the highest R^2 values, indicating higher degree of correlation, are calculated for the linking functions that predict the volume of TDS > 500 mg/l. A lower correlation is associated with the linking functions for estimating the volume of aquifers contaminated with As. We also built scatter plots of the responses estimated with the linking function and those of the complex model (Figure 6). In general, the cloud of points is distributed along a $y = x$ line, showing the accuracy of the fitting. The best correlation is for smaller simulations times.
- Step 6. Application of the linking function to approximate the output of the complex model from the outputs of two simple models. For the NRAP evaluations, the outputs

from the simple models will be the volumes of impacted aquifer estimated with two second-generation ROMs developed on the basis of the hydrogeological and geochemical conditions of the High Plain Aquifer (Section 2). In practice, once the outputs from ROMs for a particular objective variable (e.g., TDS, pH, As, Cd, or Pb) and simulation time are calculated, the estimated values will be directly used as input in the linking functions to estimate the impacted aquifer volume, considering both the hydrogeological and geochemical uncertainties. This “final” volume is expected to represent a reasonable approximation of the output from a time consuming and computationally expensive computer model. The work flow for the application of the linking function methodology is shown in Figure 27.

Table 6: Input parameters of the numerical experiment.

Parameter	Range (min – max)	Model
Porosity (unit 1)	0.25 – 0.50	Complex model, Simple model 2-D
Porosity (unit 2)	0.33 – 0.60	Complex model, Simple model 2-D
Rock density (unit 1)	2400 - 2800 kg/m ³	Complex model, Simple model 2-D
Rock density (unit 2)	2400 - 2800 kg/m ³	Complex model, Simple model 2-D
Permeability (unit 1)	-13.5 – -10.5* log(m ²)	Complex model, Simple model 2-D
Permeability (unit 2)	-15.0 – -18.0* log(m ²)	Complex model, Simple model 2-D
Van Genuchten parameter m (unit 1)	0.52 – 0.79	Complex model, Simple model 2-D
Van Genuchten parameter m (unit 2)	0.06 – 0.32	Complex model, Simple model 2-D
Van Genuchten parameter alpha (unit 1)	-4.69 – -3.81* log(m ⁻¹)	Complex model, Simple model 2-D
Van Genuchten parameter alpha (unit 2)	-5.50 – -4.14* log(m ⁻¹)	Complex model, Simple model 2-D
CO ₂ leakage rate scaling parameter ¹	0.1 – 1.0	Complex model, Simple model 2-D
Brine leakage rate scaling parameter ²	0.1 – 1.0	Complex model, Simple model 2-D
Chloride concentration in brine	-2.0 – 1.0* log(mol/L)	Complex model, Simple model 1-D
Arsenic concentration in brine	-9.0 – -5.0* log(mol/L)	Complex model, Simple model 1-D
Calcite initial volume fraction	0 – 0.2	Complex model, Simple model 1-D
Sorption scaling parameter ³	-2.0 – 2.0*	Complex model, Simple model 1-D

*Indicates log₁₀ values.

¹This factor was applied to the maximum CO₂ leakage rate.

²This factor was applied to the maximum brine leakage rate.

³This factor was applied to the adsorption capacity of different mineral phases.

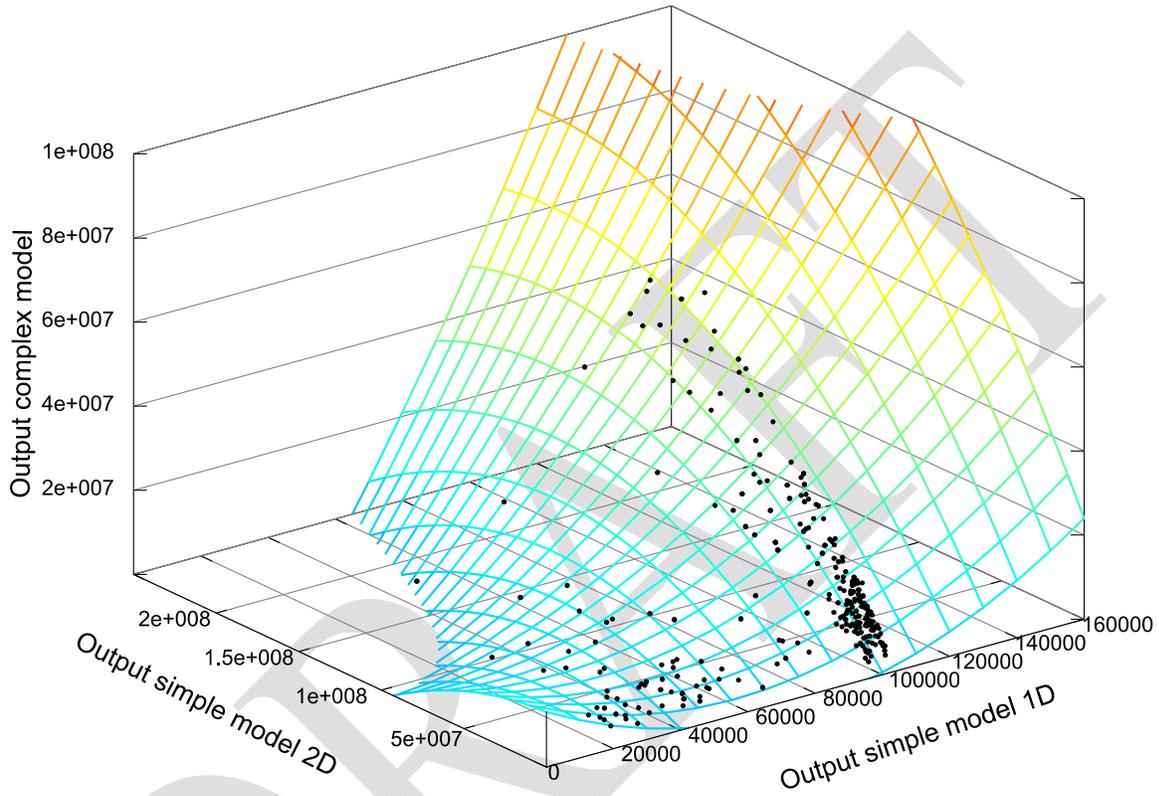


Figure 25. Second order polynomial linking function for estimating the volume (m³) of pH < 6.5 after 180 days of leakage. Points represent the calculated responses from the simple and complex models.

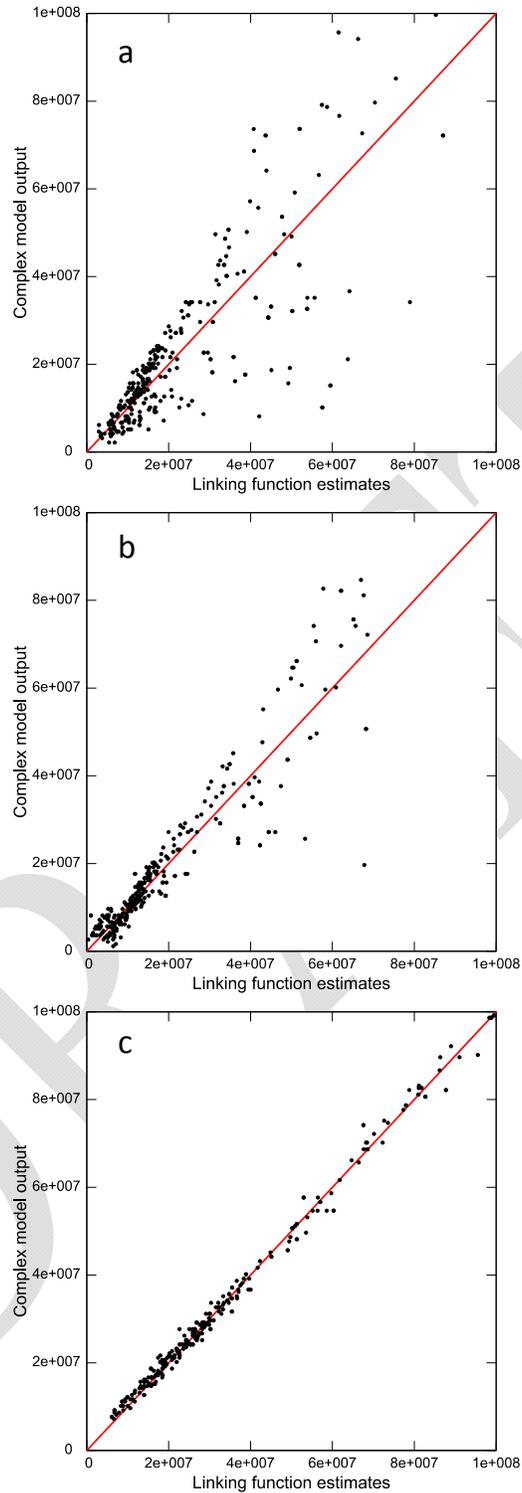


Figure 26. Comparison between complex model responses and those of the linking functions. The simulation time is 200 years. As (a); pH (b); TDS (c). The solid line represents output volumes that linking function estimates are the same as complex model output.

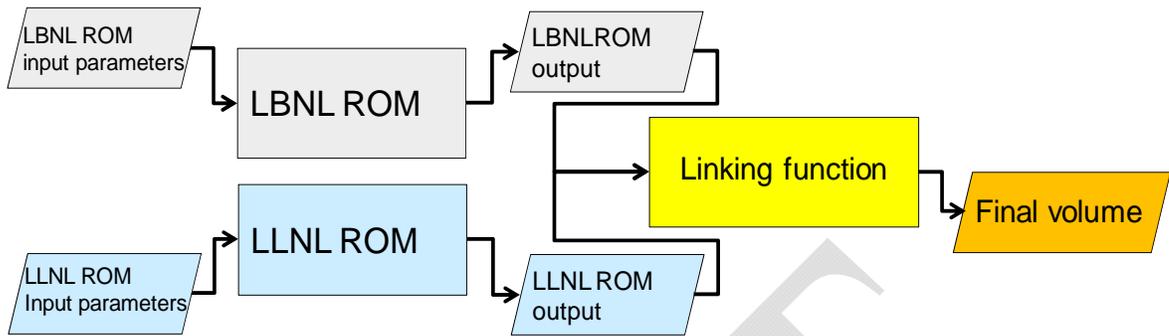


Figure 27. Flow chart for applying the linking function approach.

DRAFT

4. SUMMARY AND DISCUSSION

This report documents two activities carried out at LBNL by the NRAP Groundwater Protection Working Group: the second-generation ROM and the development and application of a linking approach. The description provided here regarding how the second-generation ROM and linking function were developed is intended to assist the NRAP team in applying this method for generating risk profiles.

In the first-generation profiles, pH and TDS are used as risk-monitoring metrics to measure the impact of potential CO₂ and brine leakage on the groundwater quality in a shallow aquifer. The second-generation risk profile includes trace metals such as Pb, As, and Cd. The spatial and temporal evolution of these metals are strongly affected by coupled flow, transport, and chemical processes. Developing one ROM directly from a model that considers all the physical and chemical processes and spatial heterogeneity remains very challenging. Until a ROM for such a complex model can be developed in a computationally efficient manner, the development of several ROMs based on simplified models and the linkage of these ROMs with a linking function would seem to be an attractive alternative.

We have developed two separate ROMs using physical and chemical properties based on the High Plains Aquifer. The first ROM, called the hydrological ROM, is derived from a model that considers 3-D heterogeneous flow and transport fields and multiple leakage wells, but is simplified in terms of chemical reactions. It considers the uncertainties of model input parameters related to flow, transport, and leakage scenarios. The second ROM, developed by LBNL, was derived from a 1-D model considering simplified flow and transport parameterization, but complex geochemical reactions. This ROM incorporates the uncertainties of model input parameters related to chemical reactions. A linking function method was applied to combine the outputs estimated by the hydrological ROM with the outputs from the LBNL ROM, allowing the combined analysis of uncertainties related to hydrological and geochemical parameters.

The development of a linking function was accomplished by: (1) the establishment of two simple models— a 1-D model with homogenous flow and transport field and single leakage point, but including as many chemical reactions as possible—and a 2-D model considering aquifer heterogeneity but no reactions; (2) the development of a complex model that incorporates all the parameters and physical and chemical processes of both simple models; (3) multiple runs of both simple and complex models; and (4) estimation of the linking functions based on those runs. The latter are expressed as a series of second order polynomials, through a least-squares regression analysis in which the outputs from the simple 1-D and 2-D models are used as independent variables to fit the outputs from the complex model. It should be noted that the complex model used here to “train” the linking function is only 2-D. Therefore, the linking function can only be applied to the hydrological ROM with the assumption that the linking function remains valid, even though the model domain increases from 2-D to 3-D.

5. REFERENCES

- Aanonsen, S. I.; Eydinov, D. A multiscale method for distributed parameter estimation with application to reservoir history matching. *Computational Geosciences* **2006**, *10*, 97-117.
- Aiuppa, A.; Federico, C.; Allard, P.; Gurrieri, S.; Valenza, M. Trace metal modeling of groundwater-gas-rock inter-actions in a volcanic aquifer: Mount Vesuvius, Southern Italy. *Chemical Geology* **2005**, *216*, 289-311.
- Anitescu, G.; Tavlarides, L. L. Supercritical extraction of contaminants from soils and sediments. *J. of Supercritical Fluids* **2006**, *38*, 167-180.
- Bacon, D. H. *Reduced Order Model for the Geochemical Impacts of Carbon Dioxide, Brine and Trace Metal Leakage into an Unconfined, Oxidizing Carbonate Aquifer, Version 2.0*; PNNL-22049; Pacific Northwest National Laboratory: Richland, Washington, 2012.
- Bear J. *Dynamics of Fluids in Porous Media*, 1972, American Elsevier Publishing Company.
- Berci, M.; Gaskell, P. H.; Hewson, R. W.; Toporov, V. V. Multifidelity metamodel building as a route to aeroelastic optimization of flexible wings. Proceedings of the Institution of Mechanical Engineers, Part C: *Journal of Mechanical Engineering Science* **2011**, *225*, 2115-2137.
- Becker, M. F.; Bruce, B. W.; Pope, L. M.; Andrews, W. J. *Ground-Water Quality in the Central High Plains Aquifer, Colorado, Kansas, New Mexico, Oklahoma, and Texas, 1999*; Water-Resources Investigations Report, 02-4112; U.S. Geological Survey, 2002.
- Bowman, K. P.; Sacks, J.; Chang, Y. F. Design and analysis of numerical experiments. *Journal of Atmospheric Science* **1993**, *50*, 1267-1278.
- Carle, S. F.; Fogg, G. E. Transition probability-based indicator geostatistics, *Math. Geol.* **1996**, *28*, 453-476.
- Carle, S. F. *T-PROGS: Transition Probability Geostatistical Software, version 2.1*. University of California: Davis, California, 1999.
- Carle, S. F.; Fogg, G. E. Modeling spatial variability with one and multidimensional continuous-lag Markov chains. *Math. Geol.* **1997**, *29*, 891-918.
- Carle S. F.; LaBolle, E. M.; Weissmann, G. S.; VanBrocklin, D.; Fogg, G. E. *Geostatistical simulation of hydrostratigraphic architecture: a transition probability / Markov approach*; In Concepts in Hydrogeology and Environmental Geology No. 2, SEPM Special Publication; 1998; p 147-170.
- Carroll, S. A.; Mansoor, K.; Sun, Y. Second Generation Reduced Order Model for Calculating Groundwater Impacts as a Function of pH, Total Dissolved Solids, and Trace Metal Concentration; NRAP-TRS-III-###-2013; NRAP Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2013; p XX
- Clement, T. P. Complexities in Hindcasting Models—When Should We Say Enough Is Enough? *Ground Water* **2011**, *49*, 620-629.
- Doherty, J. Modeling: Picture perfect or abstract art. *Ground Water* **2011**, *49*, 455-456.

- Doherty, J.; Christensen, S. Use of paired simple and complex models to reduce predictive bias and quantify uncertainty. *Water Resour. Res.* **2011**, *47*.
- Forrester, A. I. J.; Sobester, A.; Keane, A. J. Multi-fidelity optimization via surrogate modelling. *Proc. R. Soc. A* **2007**, *463*, 3251–3269.
- Gano, S. E.; Renaud, J. E.; Martin, J. D.; Simpson, T. W. Update strategies for kriging models used in variable fidelity optimization. *Struct. Multidiscip. Optim.* **2006**, *32*, 287–298.
- Grimm, V.; Frank, K.; Jeltsch, F.; Brandl, R.; Uchmanski, J.; Wissel, C. Pattern-oriented modelling in population ecology. *Sci. Total Environ.* **1996**, *183*, 151–166.
- Gómez-Hernández, J. J. Complexity. *Ground Water* **2006**, *44*, 782–785.
- Haitjema, H. The role of hand calculations in ground water modeling. *Ground Water* **2006**, *44*, 786–791.
- Hill, M. C. The practical use of simplicity in developing ground water models. *Ground Water* **2006**, *44*, 775 – 781.
- Kansas Geological Survey. Water Well Completion Records (WWC5) Database. www.kgs.ku.edu/Magellan/WaterWell/index.html (2012).
- Kharaka, Y. K.; Thordsen, J. J.; Kakouros, E.; Ambats, G.; Herkelrath, W. N.; Beers, S. A.; Birkholzer, J. T.; Apps, J. A.; Spycher, N. F.; Zheng, L.; Trautz, R. C.; Rauch, H. W.; Gullickson, K. Changes in the chemistry of shallow groundwater related to the 2008 injection of CO₂ at the ZERT Field Site, Bozeman, Montana. *Environmental Earth Sciences* **2010**, *60*, 273–284.
- Kharaka, Y. K.; Thordsen, J. J.; Hovorka, S. D.; Nance, H. S.; Cole, D. R.; Phelps, T. J.; Knauss, K.G. Potential environmental issues of CO₂ storage in deep saline aquifers: Geochemical results from the Frio - I brine pilot test, Texas, USA. *Applied Geochem.* **2009**, *24*, 1106–1112 .
- Kennedy, M. C.; O’Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **2000**, *87*, 1–13.
- Lee, D.B. Requiem for large-scale models. *J. Am. I. Planners* **1973**, *39*, 163–178.
- Little, M. G.; Jackson, R. B. Potential impacts of leakage from deep CO₂ geosequestration on overlying freshwater aquifers. *Environmental Science & Technology* **2010**, *44*, 9225–9232.
- Lødøen, O. P.; Tjelmeland, H. Bayesian calibration of hydrocarbon reservoir models using an approximate reservoir simulator in the prior specification. *Statis. Modell.* **2010**, *10*, 89–111.
- Logan, J.A. In defense of big ugly models. *Am. Entomol.* **1994**, *40*, 202–207.
- Madsen, J. I.; Langthjem, M. Multifidelity response surface approximations for the optimum design of diffuser flows. *Optim. Eng.* **2001**, *2*, 453–468.
- Razavi, S.; Tolson, B. A.; Burn, D. H. Review of surrogate modeling in water resources. *Water Resour. Res.* **2012**, *48*..

- Sacks, J.; Welch, W. J.; Mitchell, T. J.; Wynn, H. P. Design and analysis of computer experiments. *Statist. Sci.* **1989**, *4*, 409–423.
- Saltelli A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis: The Primer*; John Wiley and Sons, 2008; pp 292.
- Scheidt, C.; Caers, J.; Chen, Y.; Durlafsky, L. J. A multi-resolution workflow to generate high-resolution models constrained to dynamic data. *Comput. Geosci.* **2011**, *15*, 545–563.
- Shukhman, B. Generation of quasi-random (LP τ) vectors for parallel computation. *Computer Physics Communications*, 1994, 78(3), 279–286.
- Sun, G.; Li, G.; Stone, M.; Li, Q. A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials. *Comput. Mater. Sci.* **2010**, *49*, 500–511.
- Tong, C. Self-validated variance-based methods for sensitivity analysis of model outputs. *Reliab. Eng. Syst. Safe.* 2010, *95*, 301–309.
- Trautz, R. C.; Pugh, J. D.; Varadharajan, C.; Zheng, L.; Bianchi, M.; Nico, P. S.; Spycher, N. F.; Newell, D. L.; Esposito, R. A.; Wu, Y.; Dafflon, B.; Hubbard, S. S.; Birkholzer, J. T. Effect of Dissolved CO₂ on a Shallow Groundwater System: A Controlled Release Field Experiment. *Environmental Science & Technology* **2012**, *47*, 298–305.
- Van Nes, E. H. and M. Scheffer. A strategy to improve the contribution of complex simulation models to ecological theory. *Ecological Modelling* 2005, *185*(2–4), 153–164.
- Viana, F. A. C.; Steffen, V. Jr.; Butkewitsch, S.; Leal, M. D. F. Optimization of aircraft structural components by using nature-inspired algorithms and multi-fidelity approximations. *J. Global Optim.* **2009**, *45*, 427–449.
- Wilkin, R. T.; Digiulio, D. C. 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.
- Xu, T.; Spycher, N.; Sonnenthal, E.; Zheng, L.; Pruess, K. *TOUGHREACT User's Guide: A Simulation Program for Non-isothermal Multiphase Reactive Transport in variably Saturated Geologic Media, Version 2.0*. Earth Sciences Division, Lawrence Berkeley National Laboratory University of California: Berkeley, CA, 2012;. pp 240.
- Zheng, L.; Apps, J. A.; Zhang, Y.; Xu, T.; Birkholzer, J. T. On mobilization of lead and arsenic in groundwater in response to CO₂ leakage from deep geological storage. *Chemical Geology* **2009**, *268*, 281–297.
- Zheng, L.; Apps, J. A.; Spycher, N.; Birkholzer, J. T.; Kharaka, Y. K.; Thordsen, J.; Beers, S. R.; Herkelrath, W. N.; Kakouros, E.; Trautz, R. C. Geochemical modeling of changes in shallow groundwater chemistry observed during the MSU-ZERT CO₂ injection experiment. *International Journal of Greenhouse Gas Control* **2012a**, *7*, 202–217.
- Zheng, L.; Spycher, N.; Bacon, D. *Generation II Equilibrium and Kinetic Reactions for Groundwater Reduced Order Models*; NRAP milestones 3.5.1.d and 3.5.1.e.; Lawrence Berkeley National Laboratory: Berkeley, CA, 2012b.

- Zheng, L.; Spycher, N.; Bacon, D. *The Surface Complexation Reactions for Trace Elements on Goethite, HFO, Illite, Kaolinite, Montmorillonite and Calcite*; NRAP milestones 3.5.1.d and 3.5.1.e.; Lawrence Berkeley National Laboratory: Berkeley, CA, 2012c.
- Zheng, L.; Spycher, N.; Birkholzer, J.; Xu, T.; Apps, J.; Kharaka, Y. *Modeling studies on the transport of benzene and H₂S in CO₂-water systems*; Technical Report; LBNL-4339E; Lawrence Berkeley National Laboratory: Berkeley, CA, 2010.
- Zheng, L.; Spycher, N.; Birkholzer, J.; Xu, T.; Apps, J.; Kharaka, Y. On modeling the potential impacts of CO₂ sequestration on shallow groundwater: transport of organics and co-injected H₂S by supercritical CO₂ to shallow aquifers. *International Journal of Greenhouse Gas Control* **2013**, in press.

DRAFT

APPENDIX A: LINKING FUNCTIONS PARAMETERS

The notation used for the coefficients of the second order polynomials are as follows:

$$V = x_1 V_{simple1D} + x_2 V_{simple2D} + x_1 x_1 (V_{simple1D})^2 + x_1 x_2 V_{simple1D} V_{simple2D} + x_2 x_2 (V_{simple2D})^2$$

Linking functions parameters for estimating the volume of aquifer (m³) with AsO₃ > MCL

Time	10	20	30	40	50	60	70	80	90	100
x1	NaN	1.12E+02	1.02E+02	1.91E+02	1.46E+02	9.57E+01	1.07E+02	7.73E+01	5.39E+01	3.26E+01
x2	NaN	4.20E-01	2.10E-01	-6.03E-02	-1.15E-01	-1.14E-01	-1.35E-01	-1.03E-01	-6.73E-02	-3.65E-02
x1x1	NaN	1.06E-02	-3.41E-02	-1.93E-02	-7.16E-03	-3.42E-03	-3.00E-03	-1.89E-03	-1.14E-03	-6.96E-04
x1x2	NaN	5.78E-05	8.73E-05	6.24E-05	3.45E-05	2.44E-05	1.98E-05	1.62E-05	1.36E-05	1.18E-05
x2x2	NaN	-2.89E-10	-4.95E-09	-2.74E-09	-1.57E-09	-2.19E-09	-2.17E-09	-2.46E-09	-2.52E-09	-2.56E-09

Time	110	120	130	140	150	160	170	180	190	200
x1	3.610E+01	5.050E+01	7.851E+01	1.084E+02	1.252E+02	1.326E+02	1.370E+02	1.446E+02	1.464E+02	1.522E+02
x2	-1.103E-02	-2.118E-02	-3.186E-02	-3.257E-02	-9.299E-03	1.255E-02	1.503E-02	2.021E-02	2.614E-02	2.650E-02
x1x1	-6.603E-04	-7.231E-04	-9.747E-04	-1.196E-03	-1.262E-03	-1.213E-03	-1.143E-03	-1.123E-03	-1.070E-03	-1.046E-03
x1x2	9.909E-06	8.745E-06	7.922E-06	7.094E-06	6.177E-06	5.396E-06	4.832E-06	4.402E-06	4.085E-06	3.797E-06
x2x2	-2.266E-09	-1.964E-09	-1.780E-09	-1.573E-09	-1.413E-09	-1.272E-09	-1.099E-09	-9.969E-10	-9.242E-10	-8.582E-10

Linking functions parameters for estimating the volume of aquifer (m³) with pH < 6.5

Time	10	20	30	40	50	60	70	80	90	100
x1	NaN	NaN	1.464E+02	-1.453E+01	-1.305E+02	7.326E+01	7.407E+01	5.471E+01	1.822E+01	-1.718E+01
x2	NaN	NaN	1.006E-01	-5.709E-02	-1.461E-01	-8.411E-02	-1.281E-01	-1.088E-01	-8.083E-02	-4.080E-02
x1x1	NaN	NaN	-7.197E-02	-7.046E-03	1.922E-03	-2.229E-03	-1.736E-03	-9.980E-04	-2.435E-04	3.084E-04
x1x2	NaN	NaN	1.050E-04	5.589E-05	3.356E-05	2.005E-05	1.653E-05	1.320E-05	1.145E-05	9.916E-06
x2x2	NaN	NaN	-1.549E-09	-5.363E-10	-1.318E-09	-1.363E-09	-1.212E-09	-1.401E-09	-1.710E-09	-2.038E-09

Reduced Order Models for Prediction of Groundwater Quality Impacts from CO₂ and Brine Leakage—Application to the High Plains Aquifer

Time	110	120	130	140	150	160	170	180	190	200
x1	-2.489E+01	-3.918E+01	-5.831E+01	-5.849E+01	-5.164E+01	-4.337E+01	-3.952E+01	-3.381E+01	-2.806E+01	-3.094E+01
x2	-3.287E-02	-2.522E-02	-1.529E-03	1.157E-02	1.702E-02	1.312E-02	1.425E-02	1.212E-02	6.895E-03	1.150E-02
x1x1	3.867E-04	5.725E-04	7.718E-04	7.109E-04	5.921E-04	5.011E-04	4.345E-04	3.622E-04	3.027E-04	3.015E-04
x1x2	8.645E-06	7.650E-06	6.565E-06	5.875E-06	5.306E-06	4.819E-06	4.436E-06	4.153E-06	3.900E-06	3.651E-06
x2x2	-1.881E-09	-1.767E-09	-1.626E-09	-1.544E-09	-1.414E-09	-1.234E-09	-1.126E-09	-1.041E-09	-9.325E-10	-8.853E-10

Linking functions parameters for estimating the volume of aquifer (m³) with TDS > 500 mg/l

Time	10	20	30	40	50	60	70	80	90	100
x1	-1.875E+02	1.989E+01	1.825E+02	1.062E+02	1.501E+02	2.818E+02	2.836E+02	3.976E+02	2.378E+02	3.843E+02
x2	1.325E+00	1.205E+00	7.389E-01	4.377E-02	-1.017E+00	-2.214E+00	-2.419E+00	-3.117E+00	-2.008E+00	-3.226E+00
x1x1	8.935E-02	1.251E-02	-2.118E-02	-7.130E-03	-5.030E-03	-7.004E-03	-5.611E-03	-6.606E-03	-3.565E-03	-5.199E-03
x1x2	-2.315E-05	-9.244E-06	5.607E-05	7.979E-05	8.139E-05	8.692E-05	7.247E-05	7.170E-05	4.737E-05	5.886E-05
x2x2	-2.022E-08	-5.443E-09	-8.938E-10	-1.762E-09	6.491E-10	1.486E-09	2.016E-09	1.807E-09	5.646E-10	3.893E-10

Time	10	20	30	40	50	60	70	80	90	100
x1	2.551E+02	2.551E+02	1.837E+02	2.663E+01	1.826E+02	2.928E+02	5.500E+02	1.646E+02	1.750E+02	6.599E+02
x2	-3.909E+00	-3.909E+00	-3.776E+00	-3.557E+00	-4.072E+00	-4.874E+00	-7.927E+00	-8.415E+00	-9.986E+00	-1.428E+01
x1x1	-2.939E-03	-2.939E-03	-1.963E-03	-3.435E-04	-1.725E-03	-2.550E-03	-4.410E-03	-1.253E-03	-1.237E-03	-4.420E-03
x1x2	5.663E-05	5.663E-05	5.094E-05	4.552E-05	4.734E-05	5.132E-05	7.236E-05	7.182E-05	7.906E-05	1.036E-04
x2x2	-4.756E-10	-4.756E-10	-3.923E-10	-4.749E-10	-4.097E-10	-3.588E-10	-1.481E-10	-1.373E-10	-2.948E-12	1.149E-10



NRAP is an initiative within DOE's Office of Fossil Energy and is led by the National Energy Technology Laboratory (NETL). It is a multi-national-lab effort that leverages broad technical capabilities across the DOE complex to develop an integrated science base that can be applied to risk assessment for long-term storage of carbon dioxide (CO₂). NRAP involves five DOE national laboratories: NETL Regional University Alliance (NETL-RUA), Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), Los Alamos National Laboratory (LANL), and Pacific Northwest National Laboratory (PNNL). The NETL-RUA is an applied research collaboration that combines NETL's energy research expertise in the Office of Research and Development (ORD) with the broad capabilities of five nationally recognized, regional universities—Carnegie Mellon University (CMU), The Pennsylvania State University (PSU), University of Pittsburgh (Pitt), Virginia Polytechnic Institute and State University (VT), and West Virginia University (WVU)—and the engineering and construction expertise of an industry partner (URS Corporation).

Technical Leadership Team

Jens Birkholzer

LBNL Technical Coordinator
Lawrence Berkeley National Laboratory
Berkeley, CA

Grant Bromhal

NETL Technical Coordinator
Lead, Reservoir Performance Working Group
Office of Research and Development
National Energy Technology Laboratory
Morgantown, WV

Chris Brown

PNNL Technical Coordinator
Pacific Northwest National Laboratory
Richmond, WA

Susan Carroll

LLNL Technical Coordinator
Lawrence Livermore National Laboratory
Livermore, CA

Josh White

Lead, Induced Seismicity Working Group
Lawrence Livermore National Laboratory
Livermore, CA

Diana Bacon

Lead, Groundwater Protection Working Group
Pacific Northwest National Laboratory
Richmond, WA

Tom Daley

Lead, Strategic Monitoring Working Group
Lawrence Berkeley National Laboratory
Berkeley, CA

George Guthrie

Technical Director, NRAP
Office of Research and Development
National Energy Technology Laboratory
Pittsburgh, PA

Rajesh Pawar

LANL Technical Coordinator
Lead, Systems/Risk Modeling Working Group
Los Alamos National Laboratory
Los Alamos, NM

Tom Richard

Deputy Technical Director, NRAP
The Pennsylvania State University
NETL-Regional University Alliance
State College, PA

Nik Huerta

Lead, Migration Pathways Working Group
Office of Research and Development
National Energy Technology Laboratory
Pittsburgh, PA



Sean Plasynski
Deputy Director
Strategic Center for Coal
National Energy Technology Laboratory
U.S. Department of Energy

Jared Ciferno
Director
Office of Coal and Power R&D
National Energy Technology Laboratory
U.S. Department of Energy

Susan Maley
Technology Manager
Crosscutting Research
National Energy Technology Laboratory
U.S. Department of Energy

Mark Ackiewicz
Director
Division of Cross-cutting Research
Office of Fossil Energy
U.S. Department of Energy

NRAP Executive Committee

Cynthia Powell
Director
Office of Research and Development
National Energy Technology Laboratory
U.S. Department of Energy

Alain Bonneville
Laboratory Fellow
Pacific Northwest National Laboratory

Donald DePaolo
Chair, NRAP Executive Committee
Associate Laboratory Director
Energy and Environmental Sciences
Lawrence Berkeley National Laboratory

Melissa Fox
Program Manager
Applied Energy Programs
Los Alamos National Laboratory

Roger Aines
Chief Energy Technologist
Lawrence Livermore National
Laboratory

George Guthrie
Technical Director, NRAP
Office of Research and Development
National Energy Technology Laboratory

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.