

# LA-UR-13-22306

Approved for public release; distribution is unlimited.

Title: Generation II Reduced Order Models for Groundwater Impacts due to CO<sub>2</sub> and Brine leakage

Author(s): Keating, Elizabeth

Intended for: Report



**Disclaimer:**

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

**Generation II Reduced Order Models  
for Groundwater Impacts due to CO<sub>2</sub> and Brine leakage**

Elizabeth Keating

Zhenxue Dai

Hari Viswanathan

Los Alamos National Laboratory  
March 31, 2013

This work was funded by the Department of Energy's  
National Risk Assessment Program



## 1. Executive summary

This report describes the development of reduced-order models (ROMs) describing time-evolution of shallow groundwater impacts due to CO<sub>2</sub> leaks. These ROMs were designed to be subcomponents of a systems-level model being developed by the National Risk Assessment Project. Our work is meant to represent a particular class of groundwater systems: unconfined carbonate aquifers. While actual response of a particular aquifer to a CO<sub>2</sub> leak would be affected by complex hydrogeologic and geochemical processes, the analyses presented in this report considered complex hydrogeology but only simple geochemistry. Complimentary studies were performed at PNNL, emphasizing complex geochemistry and relatively simple (2-D) hydrogeology. Separation of these tasks was necessary due to the computational burden of simulating all the necessary processes in a single model. Similar studies are being performed at LBNL and LLNL to represent a class of confined, sandstone aquifers

These reduced-order models (“Generation 2.0 and 2.1 ROMs”) predict groundwater impacts due to CO<sub>2</sub> and brine entering a shallow aquifer from leaky wells. Groundwater impacts refer to the volume and dimensions of plumes of low pH and/or elevated TDS and trace metals, as well as the flux of CO<sub>2</sub> across the water table. The work builds on “Generation 1.0” analyses by adding two additional factors: 1) consideration of multiple leaky wells, and 2) addition of trace metal contamination. In addition, there are other improvements over the Generation I work, primarily in the development of the process models underlying the ROMs.

As in the previous Generation I work, the shallow aquifer is meant to represent a hypothetical carbonate aquifer with hydrochemical characteristics similar to the unconfined portion of the Edwards Aquifer in Texas. The aquifer model considers buffering reactions due to carbonate dissolution, but does not include any reactions that affect trace metal concentrations. The purpose of the trace metal predictions is to demonstrate possible contamination due to conservative transport of trace metals originating in the leaking brine.

A set of multiple-wellbore leakage scenarios was developed using the Generation I leaky wellbore ROM, pressure and CO<sub>2</sub> saturation profiles derived from Kimberlina reservoir simulations (Bromhal, 2012). CO<sub>2</sub> leak rates typically ranged between 0.1 and 1 g/s and brine leak rates typically ranged between 0.01 and 0.1 for this set. This set was supplemented with a small number of additional scenarios with higher peak leakage rates (up to 320 g/s CO<sub>2</sub> and 50 g/s brine) to extend the capability of the resulting ROM.

The ROMs were developed in three steps: first, running a series of multi-phase reactive transport models to span the range of uncertain inputs, and then post-

processing the results of these models to develop a response surface (RS) for groundwater impacts due to a single leaky well. Global sensitivity analysis methods were used to identify the most important variables. Generally speaking, the most important variables were CO<sub>2</sub> and brine leakage rates. Finally, the response surfaces were integrated into a reduced order model (ROM) that calculates total response to multiple leaky wells.

The resulting ROM accepts the following input from the systems model: time (years), and number of leaking wells (N). For each leaking well, the following input is required: xy location of the well, instantaneous CO<sub>2</sub> and brine leakage rate (kg/s), and cumulative mass of leaked CO<sub>2</sub> and brine (kg). The following results are returned: total CO<sub>2</sub> flux to the water table, volume of aquifer exceeding critical threshold values (C\*) for five metrics: pH, TDS, [Cd], [As], and [Pb], and the x and y dimensions (m) of the pH plume. Critical threshold values were assumed to be MCLs for the Generation 2.0 ROMs, but were lowered significantly to approximate 'background values' for Generation 2.1 ROMs (see Table 1).

Table 1. Initial conditions and two sets of critical threshold values (C\*)

		Critical threshold values	
	Initial aquifer value	MCL	Background
pH	6.9	6.5	6.6
TDS (mg/l)	330	500	420
As (µg/l)	0.31	10	0.55
Pb (µg/l)	0.064	15	0.15
Cd (µg/l)	0.0	5	0.04

Summary statistics of inputs and outputs for these ROMs are shown in Table 2. The largest plume volumes were for depressed pH compared to the background levels (maximum for all simulations: 5.2E7 m<sup>3</sup>). This is roughly a plume 400 X 800 X 150 m in dimension. The median plume size was much smaller: 3.02E6 m<sup>3</sup>, roughly a plume 100 X 200 X 150 m in dimension. Less than 1% of the runs had 'no impact', meaning vol of pH plume ~ 0 at 200 years. TDS plume volumes tended to be slightly smaller than pH volumes. For the trace metals considered (As, Pb, and Cd), the plumes were very small (median values ~500 m<sup>3</sup>) when calculated using the MCL as the critical threshold. When the background concentrations were used, the trace metal plume volumes were about 20 – 50% the size of the TDS plume volumes.

The plumes generally moved vertically to the water table with less than 1km migration in the direction of groundwater flow over 200 years (maximum downstream migration distance = 822 meters). Once the plume reaches the water table, the flux rate of CO<sub>2</sub> across the water table rapidly equilibrates to match the flux of CO<sub>2</sub> at the wellbore.

Table 2. Input and output parameters for the reduced order models

ROM input parameters

	Parameter	Symbol	Min	Max	Median
Aquifer characteristic	Variance	$\sigma$	0.017	0.79	0.34
	Correlation length (km)	$\lambda$	1	3.95	2.75
	Anisotropy	K <sub>xy</sub> /K <sub>z</sub>	1.1	49.1	24.4
	Mean Aquifer permeability (log <sub>10</sub> m <sup>2</sup> )	k	-13.5	-10.6	-12.1
	Mean aquifer porosity	$\Phi$	0.05	0.34	0.13
CO <sub>2</sub> leak	CO <sub>2</sub> leak rate (g/s)	C1	0.00001	311	0.245
	Cumulative CO <sub>2</sub> mass (kton)	C2	0.00124	1840	1.41
Brine leak	Brine leak rate (g/s)	B1	0.0018	36	0.0102
	Brine mass (kton)	B2	0.0112	291	0.068
	[Cl] (mol/l)	[Cl]	0.1	6.7	2.1

Output metrics (at 200 years)

			C* MCL (n=161)			C* Threshold (n=158)	
		Min	Max	Median	Min	Max	Median
Flux to water table	(g/s)	0	306	4.42E-4	0	306	2.025E-3
TDS	Volume (m <sup>3</sup> )	1.66E+04	3.03E+07	4.76E+05	3.47E+04	3.65E+07	6.52E+05
As		0	2.47E+06	5.06E+02	5.06E+02	1.68E+07	1.49E+05
Pb		0	2.48E+06	5.18E+02	1.14E+04	2.70E+07	3.89E+05

Cd		0	2.05E+05	0	2.44E+02	1.41E+07	1.06E+05
pH	Volume (m <sup>3</sup> )	0	4.56E+07	3.02E+06	0.00E+00	5.20E+07	3.55E+06
	$\Delta x$	0	184	56	0	526	64
	$\Delta y$	0	716	184	0	822	214

## 2. Shallow aquifer model simulations

Most aspects of the shallow groundwater model simulations were identical to Generation I (Dai et al., 2011). Simulations were performed with the code FEHM (Finite-Heat-and-Mass-Transfer) code (Zyvoloski, 1997). Lateral boundaries were hydrostatic, the water table was assumed to have a gradient of 7.5 Pa/m, and the aquifer is 150m thick. Assuming the median aquifer permeability and porosity (Table 2) this corresponds to a pore water velocity of 1.3 m/year. In other words, a conservative tracer would be expected to travel ~184m in 200 years. Stochastic fields of heterogeneous permeability were generated using the pilot point method and random Gaussian interpolation. All nodes were assumed to have anisotropic intrinsic permeability.

Several improvements were made to the Generation I shallow aquifer model. The first was to decrease the grid size near the leaky well. The Generation I model used a constant grid spacing  $dx=dy=100m$ ;  $dz=10m$ ., and recent inspection of those results showed a significant grid size effect. If the grid cell near the injection well is sufficiently large, the rate of dissolution into the groundwater can accommodate the CO<sub>2</sub> leak rate and free (gas) CO<sub>2</sub> phase never forms. No significant buoyancy forces develop, and the plume slowly grows and moves downgradient. Alternatively, if the grid cell near the injection well is small, gas phase develops, buoyancy forces drive the plume upward, and CO<sub>2</sub> flux to the water table is enhanced. To correctly capture these physics, we developed a computational mesh with a variable grid spacing (the  $dx$ ,  $dy$ ,  $dz$  near the well are 9 m, 9m, and 8m respectively, and then they increase gradually to 200m, 300m and 20m at the far field). The size of the model domain was set to 8km X 5km X 150m. Using the range of model parameters described below, plumes never interacted with lateral model boundaries.

Generation I models assumed linear relative permeability and capillary pressure relationships. The ROMs derived for Generation II were based on simulations with more realistic relative permeability and capillary pressure relationships (Brooks-Corey). Linear models were also used for comparison.

Generation II models allowed porosity to vary spatially along with permeability. The following equation was used (Bernabe et al., 2003 and Deng et al., 2012)

$$k = a\phi^b \quad (1)$$

Where,  $k$  is permeability,  $\phi$  is porosity,  $a$  and  $b$  are coefficients ( $a=4.84 \times 10^{-10}$  and  $b=3$ ) (Bernabe et al., 2003 and Deng et al., 2012). Table 2 summarizes the range of the 5 aquifer parameters (variance and correlation length of random permeability fields, mean permeability and porosity, and permeability anisotropy) that were varied in the process model simulations. **Figure 1** illustrates one stochastic permeability field and the computational mesh.

## 2.1. Geochemistry

**Carbonate geochemistry.** Aqueous and solid phase carbonate reactions, listed in Table 2, were simulated. These should primarily affect changes in the pH of the shallow aquifer. A potential secondary effect, that as carbonate dissolves the TDS of the water may increase, was neglected in these calculations.

Table 3. Aqueous and solid phase carbonate reactions

Carbonate aqueous equilibrium reactions:	
$\text{HCO}_3^- + \text{H}^+ = \text{H}_2\text{CO}_3$	$\log K = -6.3$
$\text{CO}_3^{2-} + 2\text{H}^+ = \text{H}_2\text{CO}_3$	$\log K = -16.6$
$\text{OH}^- + \text{H}^+ = \text{H}_2\text{O}$	$\log K = -14$
Calcite kinetic dissolution or precipitation:	
$\text{CaCO}_3(\text{s}) + 2\text{H}^+ = \text{H}_2\text{CO}_3 + \text{Ca}^{++}$	

**Sodium, chloride, and trace metals.** The leaky brine was assumed to have a constant concentration of  $[\text{Cl}]$ , considered a random variable (shown in Table 2). Conservative Cl transport was simulated using the advection/dispersion equation. The aquifer was assumed to have a dispersion coefficient of 500 m. The TDS of the groundwater (mg/l) was calculated from  $[\text{Cl}]$  (moles/l) by the following equation:  $\text{TDS} = \text{TDS}_0 + [\text{Cl}] * (35.45 + 22.99) * 1000$ .  $\text{TDS}_0$  is the assumed initial TDS of the aquifer (347.22 mg/l).

Trace metal transport was assumed to be conservative, and so was calculated at the end of the simulations based on a presumed constant molar ratio between  $[\text{Cl}]$  and the three trace metals. These ratios were derived from a brine chemistry data analysis (Hakala and Lopano, 2012). They are tabulated in Table 4.

Table 4. Assumed molar ratios between  $[\text{Cl}]$  and trace metals

$[\text{As}]/[\text{Cl}]$	1.71863E-7
$[\text{Pb}]/[\text{Cl}]$	9.66456E-8
$[\text{Cd}]/[\text{Cl}]$	1.71863E-8



## 2.2. Leakage scenarios

Ultimately, the groundwater impact ROM will not be required to calculate spatial location of leaky wells, or their leakage rates. However, the ROM must accommodate a range of leak rates as inputs and so it was necessary to generate an ensemble of leakage scenarios to train the ROM. To do so, we used the following process.

**Location of leaky wells.** The spatial distribution and depths of possible leaky wells was derived from a database of wells at in the High Plains aquifer. Of these 128, the average inter-well spacing is ~1600m. Five percent of the wells are located within 150m of another well. Generally speaking, if plume sizes tend to be smaller than the spacing between leaky wells, the principle of superposition can be used to calculate total volume of contaminated water.

Because of the requirements for small grid cells near the leaky well (described above) simulating a spatially-explicit multiple-leaky well scenario would require a very large number of grid cells. If, in fact, plumes from different leaky wells do not overlap, utilizing a high-resolution grid to simulate each leak and then applying the superposition principle would be more accurate than simultaneously simulating multiple leaks and necessarily using larger grid cells near each leak. Using the high-resolution grid and a single leaky well, we simulated a large range of leakage rates. We determined that, with very few exceptions, the maximum dimensions of the simulated plumes were smaller than the average well-spacing in the 'leaky well dataset'. **Figure 2** shows the well locations and the size of the largest pH plume simulated. Therefore, we chose to proceed with the superposition principle to obtain the best possible accuracy. We added an internal check to the ROM to ensure that the size of the plumes never exceeds the spacing between two leaky wells. If that criteria is exceeded, the two wells are combined into a single source with a combined leak rate.

**Leak rates.** The generation I leaky wellbore ROM, driven by pressures and saturations predicted by the Kimberlina simulations, were used to establish a likely range of CO<sub>2</sub> and brine leakage. To create a range of input parameters for the leaky wellbore ROM, the following parameters were varied: leaky wellbore permeability ( $P_w$ ), reservoir over-pressurization ( $dP_{max}$ ) and CO<sub>2</sub> saturation ( $S$ ) at the injection well (see Table 5). An empirical relationship was developed to calculate the variation of  $dP$  and  $S$  with time and with distance from the injector well, using the results of 47 stochastic simulations of the Kimberlina reservoir (ref). This relationship provided a broad range of leakage rates, but should be tested against more recent reservoir simulations to verify consistency.



Table 5. Parameters varied in the leaky wellbore ROM to create CO<sub>2</sub> and brine leakage rates

			Min	Max	Mean	Stdev	
R1	Maximum overpressurization at injection well (Mpa)	$dP_{max}$	15	40			uniform
R2	Maximum CO <sub>2</sub> saturation at injection well	$S_{max}$	0.1	1.0			uniform
R3	Permeability of leaky wellbore	$P_w$			-11	1.	normal

CO<sub>2</sub> leak rates typically ranged between 0.1 and 1 g/s and brine leak rates typically ranged between 0.01 and 0.1 for this set. All the scenarios share the common characteristics of fairly early rise to a peak rate (~ 50 years) which stays fairly constant afterwards. This set was supplemented with a small number of additional scenarios with higher peak leakage rates (up to 320 g/s CO<sub>2</sub> and 50 g/s brine) to extend the capability of the resulting ROM.

In **Figure 3**, the final set of CO<sub>2</sub> brine and leakage rates are summarized using percentiles. Time is relative to the initial time the CO<sub>2</sub> and/or brine begins to leak any particular well.

### 2.3. Example simulation result

To provide an example of a single process model result, here we show the simulation results for one example leakage scenario (shown in **Figure 4**), the following values of aquifer parameters: mean conductivity 3.06 m/d, porosity 0.19, and a leaky brine Cl concentration of 4.1 mol/L.

The plume evolution is shown in **Figure 5**. Plume migration is dominated by vertical flow, caused by the buoyancy of the free (gas) phase CO<sub>2</sub>. The dissolved CO<sub>2</sub> plume only moves downstream a relatively short distance. For reference, the calculated volumes of plumes exceed critical thresholds are shown in **Figure 6**, as well as the CO<sub>2</sub> flux to the water table and the x and y dimensions of the pH plume.

## 3. Monte Carlo runs

Using the ranges of parameters listed in Table 2 and the leakage scenarios summarized in **Figure 3**, we generated multiple realizations, each representing a unique combination of aquifer parameters and leakage scenarios. **Figure 7** presents bivariate plots of all the parameters used. As can be seen in this figure, most parameters are uncorrelated. One exception is permeability/porosity, which are

correlated through Equation (1). The distribution of CO<sub>2</sub> and brine leak rates appears to be log-normal. This is because most of the leakage scenarios were generated with the Gen 1 Wellbore leakage ROM (fairly small leak rates) and a few larger leak rates were added to expand the total range. The distribution of permeability values was presumed to be log-normal.

The eight output variables and two input variables (cumulative CO<sub>2</sub> and brine mass) are shown in bivariate plots in **Figure 8a** (MCLs) and **8b** (background values). For a given simulation, each time appears as a separate symbol. Several important conclusions can be drawn from these plots. Examination of the last row in **Figure 8(a)** demonstrates there is a strong relationship between the cumulative brine mass leaked (B2) and the TDS, As, Pb, and Cd plumes. The scatter demonstrates the importance of other variables, as well, but clearly brine mass is the dominant variable. Also, there is an almost perfect 1:1 relationship between Pb and As plume sizes. Finally, all the output variables are log-normally distributed, suggesting a dominant role of one or more of the log-normally distributed input parameters (K, C1, C2, B1, and B2). First, there is a 1:1 relationship between All of the output variables have a log-normal distribution. A similar story emerges from the background results (**Figure 8(b)**).

A strong bivariate relationship not included in **Figure 8** is shown in **Figure 9**. There is a nearly perfect 1:1 relationship between CO<sub>2</sub> leak rate and CO<sub>2</sub> across the water table.

The time-variation of these results are shown in **Figure 10**. The median plume volumes are relatively small (dimensions of the order of 20m X 75m). The ensemble of runs do include some volumes substantially larger, but still not exceeding 526m X 822m. This result gives some confidence in the superposition principle (see **Figure 2**). Plume volumes and dimensions generally grow monotonically through time.

## 4. Global sensitivity analysis

In order to determine the key parameters driving CO<sub>2</sub> reactive transport behavior in the Monte Carlo simulations, global sensitivity analysis techniques are used for investigating input-output sensitivities that are valid over the entire range of the parameter sampling. The Monte Carlo simulations provide 200 realizations of input transport parameter sets created with Latin Hypercube sampling for each model. Each realization is propagated through the transport simulator to yield the output response functions represented by the 8 output variables (Table 2). Global sensitivity analysis entails the comparison of the output distributions to each of the input transport parameter distributions and identifies the most sensitive parameters. We used the MARS-based parameter importance analysis (Friedman, 1991) implemented using the PSUADE code (Tong, 2011). The two sets of results for each of the two critical threshold metrics (Table 1) were analyzed separately. The results are presented in **Figure 11 and 12**.

The volume and dimensions of the pH plume (**Figure 11a-c**) are dominated by CO<sub>2</sub> leak rates and two of the aquifer parameters: porosity and mean permeability. The CO<sub>2</sub> flux rate across the water table is dominated by the CO<sub>2</sub> leak rate. For the MCL threshold, the trace metal plumes are relatively insensitive to all the aquifer parameters, but a strongly sensitive to the brine mass rate. There is someone higher sensitivity to the aquifer parameters in the threshold calculations, particularly  $\lambda$ .

There is clearly some statistical noise in these results, especially at lower sensitivities, as evidenced by the non-physical nature of some of the identified sensitivities. For example, there is no physical reason that the width of the pH plume should be sensitive to the [Cl] of the leaky brine (**Figure 11b**). Therefore, these sensitivity analysis results should be interpreted qualitatively, and very small sensitivities should be viewed with caution.

By using the Monte Carlo simulation results as the input for PSUADE (Tong, 2011), we conduct global sensitivity analysis with the MARS methods.

In order to compare the results from these two methods, we put the sensitivity indices derived from these methods together in one figure. In most cases the results from these two methods are consistent with each other (except a few cases). The results plotted in **Figures 11 and 12** shows that different output variables are sensitive to different parameters. For example, the volume of pH<6.5 is most sensitive to aquifer anisotropy factor, conductivity, porosity and cumulative leaked CO<sub>2</sub> mass, while the CO<sub>2</sub> leakage rate to the atmosphere is most sensitive to aquifer conductivity, porosity, CO<sub>2</sub> leakage rate from deep reservoir and the cumulative mass. The tracer metals As, Pb and Cd are most sensitive to Cl<sup>-</sup> concentration in the leaked brine, CO<sub>2</sub> mass and aquifer anisotropy factor, because the recharge concentrations of As, Pb and Cd in the leaked brine are positively correlated to Cl<sup>-</sup> concentration.

## 5. Response Surface Development

Using these simulation results, response surfaces (RS) were developed for a single leaky well using PSUADE (Tong, 2011). Three types of response surfaces were developed, in increasing order of complexity:

Linear response surfaces:  $O(P_k) = C_0 + \sum_{k=1}^{Np} C_k P_k$

Quadratic response surfaces:  $O(P_k) = C_0 + \sum_{k=1}^{Np} C1_k P_k + \sum_{i=1}^{Np} \sum_{k=1}^{Np} C2_{ik} P_i P_k$

Cubic response surfaces:  $O(P_k) = C_0 + \sum_{k=1}^{Np} C1_k P_k + \sum_{i=1}^{Np} \sum_{k=1}^{Np} C2_{ik} P_i P_k + \sum_{i=1}^{Np} \sum_{j=1}^{Np} \sum_{k=1}^{Np} C3_{ijk} P_i P_j P_k$

It is generally expected that the high-order models will provide a better fit to the training dataset, due to the larger number of parameters available. However, higher-order models can have lower predictive capability when used for scenarios

different than those in the training dataset. One goal of this analysis is to strike the right balance between optimizing accuracy and minimizing predictive uncertainty.

The correlation between the response surface predictions and the outputs from the ‘training data set’ are presented in Table 5. All three response surfaces were incorporated into the reduced order model provided to the systems modeling team.

Correlation coefficients for all six response surfaces are reported in Table 5. The highest correlations are for the cubic surfaces; the lowest are for the linear surfaces. Informal cross-validation studies, however, showed that the cubic response surfaces performed poorly when predicting outputs for model runs not used in training the surface. We performed a model selection analysis (presented in Appendix I) and found the quadratic response surfaces to be the best models, as measured by highest goodness-of-fit with the lowest parameter uncertainty. As shown in **Figure 13**, the quadratic RS performs very well for all metrics with the exception of pH plume volume. The largest errors for pH plume volumes are of the order of  $1E7m^3$ . These errors reflect the inherent limitation of a simplified RS to mimic the complex behavior of a multi-phase flow and transport simulation. Nevertheless, the RS does appear to capture the behavior to the first order.

Table 6. Correlation coefficients between process model results and response surface, for the two sets of threshold criteria (Table 1).

Output	MCL			Background		
	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
pH ( $m^3$ )	0.50	0.89	1.00	0.52	0.88	1.00
CO <sub>2</sub> flux(kg/s)	1.00	1.00	1.00	1.00	1.00	1.00
$\Delta x$ (m)	0.57	0.89	0.99	0.55	0.81	0.90
$\Delta y$ (m)	0.61	0.86	0.99	0.63	0.85	0.96
TDS ( $m^3$ )	0.95	0.98	1.00	0.94	0.97	1.00
As ( $m^3$ )	0.98	1.00	1.00	0.97	0.99	1.00
Pb ( $m^3$ )	0.98	1.00	1.00	0.96	0.98	1.00
Cd ( $m^3$ )	0.90	1.00	1.00	0.97	0.99	1.00

In **Figure 14** we present a comparison of time-varying process model outputs and the corresponding RS predictions. The best-fit model is the cubic model. However, as discussed in greater depth in Appendix I, the relatively small ratio between total number of model parameters and the size of the calibration dataset for the cubic model makes this RS an unreliable predictor.

## 6. Reduced order model

The final reduced order model (ROM) accepts input from the systems model and calls the quadratic response surfaces described above as a set of subroutines. The relationships between the systems model, the ROM, and the RS is illustrated in **Figure 15**. One functionality of the ROM is to translate between ‘absolute time’ (defined by the systems model as the time injection begins) and the time variable used by the RS, where the leak is assumed to reach the top of the wellbore at time=0. This is an important functionality because if the well is far from the injector there will be a significant delay between the time injection begins and the time a well begins to leak. The other functionality is to manage multiple leaky wells scenarios. For each leaky well, the ROM stores the time that brine begins to leak and that CO<sub>2</sub> begins to leak. The values of the time variables subsequently provided to the RS are relative to these breakthrough times. Additionally, the CO<sub>2</sub> fluxes, plume volumes and dimensions produced by each leaky well are summed to return a single integrated value for all leaky wells.

A parameter  $D_{min}$ , set internally by the systems model, defines the minimum distance between any two leaky wells for the summation of plume volumes to be valid. If any two wells are closer than  $D_{min}$ , their leak rates are combined into a single source before the RS call. Based on our process model results, it appears that 500m is a reasonable value for  $D_{min}$ . This assumption should be re-evaluated in the future as new results become available.

## **7. Limitations and need for additional work**

We have shown that simplified, computationally-efficient ROMs can be used to mimic the behavior of more complex multi-phase flow and transport models. The ROMs reproduce key simulation outputs very well, with the exception of one metric: pH plume volumes. In this case, the ROMs do reproduce the first-order behavior of the process model, but there is significant scatter around the 1:1 relationship.

For these results to be applicable to a wider range of scenarios, additional process model runs should be developed. It should be demonstrated that these results are applicable to other types of unconfined aquifers, such as sandstone unconfined aquifers, aquifers with other types of heterogeneity (non-gaussian), aquifers with lesser or greater thickness than 150m, and aquifers with different pre-leak hydraulic gradients than the 7.5Pa/m gradient assumed in all these simulations. Adding these additional factors in the analysis would make these simulations (and derived ROMs) much more broadly applicable.

Also, the critical threshold values (Table 1) should be added to the ROM analysis as input parameters that can take on any arbitrary value. We have begun a preliminary analysis to do so and the initial results are promising.

The leakage scenarios used to develop the process model results should be continuously re-evaluated as the leaky-wellbore simulations evolve and the systems modeling group develops more leaky wellbore scenarios. Formal cross-validation studies should be conducted each time new leaky scenarios are brought into the simulation ensembles so that we can document the predictive capability of the groundwater ROMs.

Finally, the response surface development should adapt a two-stage strategy, beginning with a screening to eliminate insensitive or non-physical parameter inputs, followed by developing the RS using only the remaining parameters. This may produce RS models with better goodness-of-fit and predictive reliability.

## 8. Appendix. RS selection criteria

The data quantity/quality and the number of parameters for inverse modeling have a large impact on the inversion results. Usually, with the same observational data, the more parameters involved in the inversion, the better the fit. Therefore, we cannot select the best response surface (RS) by just checking the goodness of fit (the values of inverse objective functions or the R-square values). To remove the effect of the number of parameters to the solutions of the inverse problems, we use three model identification criteria to assess ROM parsimony and to identify the best ROMs. These criteria are developed within the context of maximum likelihood theory (Carrera and Neuman, 1986; Samper and Neuman, 1989; Hill, 1998; Poeter and Anderson, 2005; Ye et al., 2008 and Dai et al., 2008, 2012). The model identification criteria minimize the measures of the closeness between the true (yet unknown) model and the proposed model. The first criterion is the Akaike information criterion (*AIC*) (Akaike, 1974; Poeter and Anderson, 2005).

$$AIC_k = N_z \ln \hat{\sigma}_{ML}^2 + 2P_k, \quad (1)$$

where  $N_z$  is the number of data used for solving the inverse problem;  $P_k$  is the number of estimated parameters;  $k$  indicates the  $k$ -th alternative process-conceptual model,  $k = \overline{1, K}$ ;  $N_z \ln \hat{\sigma}_{ML}^2$  is a term obtained from the unbiased least square estimator,

$$\hat{\sigma}_{ML}^2 = \frac{\Phi}{N_z} \Bigg|_{\theta_k = \hat{\theta}_k}, \quad (2)$$

where  $\Phi$  is the generalized or weighted sum of squared residuals yielding a parameter estimate  $\hat{\theta}_k$  from minimizing the least square objective function (Ye et al., 2008). The best model is that with the smallest *AIC* value. In order to correct the bias of *AIC* criterion for a small number of calibration data (e.g.,  $N_z/P_k < 40$ ), the modified Akaike information criterion (*AICc*) was developed as (Hurvich and Tsai, 1989),

$$AICc_k = N_z \ln \hat{\sigma}_{ML}^2 + 2P_k + \frac{2N_z(N_z + 1)}{N_z - P_k - 2}. \quad (3)$$

*AICc* considers the size of the calibration data set and is therefore preferred over the original *AIC*. Schwarz (1978) derived a different criterion in a Bayesian context, called Bayesian information criterion (*BIC*) as defined below,

$$BIC_k = N_z \ln \hat{\sigma}_{ML}^2 + P_k \ln N_z. \quad (4)$$

The inverse methodology applied by PSUADE (Tong, 2011) to develop the RS outputs the generalized least-squares objective function  $\Phi$ , which then is used to calculate the three criteria for all of the alternative (linear, quadratic and cubic) ROMs with thresholds and MCLs (see Tables 1 and 2). By comparing the values of the criteria for the linear, quadratic and cubic RSs, we find the cubic RSs for the 8 output variables can minimize the criteria of AIC, AICC and BIC. Therefore, we conclude that the cubic and quadratic RSs for the 8 output variables are the best models and can be used for system-level prediction.

Table 2. Model selection results for RSs using background values as thresholds

Linear		Response Surfaces						
Output	pHVol	co2 flux	$\Delta x$	$\Delta y$	TDS	As	Pb	Cd
Parameter#	12	12	12	12	12	12	12	12
R-square	0.5163	0.9985	0.5509	0.6271	0.9357	0.9708	0.9553	0.974
Objective	0.6954	0.039	0.67	0.6105	0.2536	0.171	0.2115	0.1613
AIC	-2287	-20606	-2523	-3114	-8702	-11210	-9856	-11581
AICC	4103	-14216	3867	3276	-2312	-4820	-3466	-5191
BIC	-2214	-20533	-2450	-3041	-8629	-11137	-9783	-11508
Quadratic		Response Surfaces						
Parameter#	78	78	78	78	78	78	78	78
R-square	0.8816	0.9998	0.8066	0.8451	0.9697	0.9909	0.9816	0.9923
Objective	0.3441	0.0142	0.4397	0.3935	0.1742	0.0955	0.1356	0.0878
AIC	-6629	-26896	-5070	-5776	-10960	-14783	-12550	-15319
AICC	-103	-20370	1457	750	-4434	-8257	-6024	-8793
BIC	-6156	-26423	-4597	-5303	-10487	-14310	-12077	-14846
Cubic		Response Surfaces						
Parameter#	364	364	364	364	364	364	364	364
R-square	0.9964	1	0.8994	0.9586	0.9974	0.9989	0.998	0.9992
Objective	0.0602	0.003	0.3171	0.2035	0.0513	0.0326	0.0442	0.0287
AIC	-17145	-36282	-6576	-9397	-18157	-21054	-19104	-21850
AICC	-9955	-29093	613	-2208	-10967	-13864	-11914	-14660
BIC	-14937	-34075	-4369	-7190	-15949	-18846	-16896	-19642

Table 2. Model selection results for RSs using MCL threshold values.

Linear								
Output	pHVol	co2 flux	$\Delta x$	$\Delta y$	TDS	As	Pb	Cd
Parameter#	12	12	12	12	12	12	12	12
R-square	0.5043	0.9985	0.5696	0.6094	0.9486	0.9773	0.9773	0.8963
Objective	0.704	0.039	0.656	0.6249	0.2268	0.1506	0.1507	0.3219
AIC	-2251	-20994	-2708	-3023	-9590	-12242	-12241	-7321
AICC	4259	-14484	3802	3487	-3080	-5732	-5731	-811
BIC	-2178	-20921	-2635	-2950	-9517	-12169	-12168	-7248
Quadratic								
Parameter#	78	78	78	78	78	78	78	78
R-square	0.8863	0.9998	0.89	0.8552	0.9777	0.9957	0.9957	0.9955
Objective	0.3372	0.0142	0.3316	0.3804	0.1492	0.0656	0.0654	0.0672
AIC	-6889	-27407	-6997	-6106	-12173	-17496	-17518	-17339
AICC	-243	-20761	-351	540	-5527	-10850	-10871	-10693
BIC	-6414	-26933	-6523	-5632	-11699	-17022	-17043	-16865
Cubic								
Parameter#	364	364	364	364	364	364	364	364
R-square	0.9967	1	0.9895	0.9872	0.9978	0.9999	0.9999	0.9999
Objective	0.0571	0.003	0.1026	0.1133	0.0472	0.0117	0.0121	0.0104
AIC	-17828	-36967	-14025	-13384	-19060	-28069	-27870	-28872
AICC	-10520	-29660	-6717	-6077	-11753	-20762	-20562	-21565
BIC	-15613	-34753	-11810	-11170	-16846	-25855	-25655	-26658

## 9. References

- Bernabe, Y., Mok, U., Evans, B., Permeability-porosity relationships in rock subjected to various evolution processes. *Pure and Applied Geophysics* 160, 937–960, 2003.
- Bromhal, G., Reservoir Performance Group ROM Discussion, NETL, Progress report, National Risk Assessment Partnership Workshop, January 9, 2012.
- Carrera, J. and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, 22(2), 199-210.
- Dai, Z., E. Keating, and H. Viswanathan, Impact of the leaked CO<sub>2</sub> from deep reservoirs on quality of shallow groundwater, 2011 NRAP progress report Los Alamos National Laboratory, 2011.



- Dai, Z. J. Samper, A. Wolfsberg, and D. Levitt (2008), Identification of relative conductivity models for water flow and solute transport in unsaturated bentonite, *Phys. Chem. Earth*, 33(suppl), s177-s185, doi:10.1016/j.pce.2008.10.012.
- Dai, Z., A. Wolfsberg, P. Reimus, H. Deng, E. Kwicklis, M. Ding, D. Ware and M. Ye, Identification of sorption processes and parameters for radionuclide transport in fractured rock, *J. of Hydrol.*, v 414–415, p220–230, 2012.
- Deng, H., P. Stauffer, Z. Dai, Z. Jiao, R. Surdam, Simulation of industrial-scale CO<sub>2</sub> storage: Multi-scale heterogeneity and its impacts on storage capacity, injectivity and leakage, *Int. J. of Greenhouse Gas Control*, 10 (2012) 397–418, 2012.
- Friedman, J. H. Multivariate adaptive regression splines, *Annals of Statistics*, 19.1, 1-141, 1991
- Hakala, A. and C. Lopano, Brine Chemistry Data for Generation II Risk Profile Development (Data from the Natcarb Brine Database), NETL, NRAP progress report, 2012.
- Hill, M.C., (1998), Methods and guidelines for effective model calibration, US Geological Survey, Water-Resources Investigations Report 98-4005.
- Hill, M. C., and C.R. Tiedeman (2007), Effective calibration of ground water models, with analysis of data, sensitivities, predictions, and uncertainty, 455p, John Wiley, New York.
- Poeter, E. and D. Anderson (2005), Multimodel ranking and inference in groundwater modeling, *Ground Water*, 43(4), 597-605.
- Samper, F. J. and S. P. Neuman (1989), Estimation of spatial covariance structures by adjoint state maximum likelihood cross validation: 1. Theory, *Water Resour. Res.*, 25(3), 351-362.
- Tong, C. (2011). PSUADE User's Manual, Lawrence Livermore National Laboratory, LLNL-SM-407882.
- Viswanathan H., Z. Dai, C. Lopano, E. Keating, J. A. Hakala, K. G. Scheckel, L. Zheng, G. D. Guthrie, R. Pawar, Developing a robust geochemical and reactive transport model to evaluate possible sources of arsenic at the CO<sub>2</sub> sequestration natural analog site in Chimayo, New Mexico, *Int. J. of Greenhouse Gas Control*, 10 (2012) 199–214, 2012.
- Ye, M., P. D. Meyer, and S. P. Neuman (2008), On model selection criteria in multimodel analysis, *Water Resour. Res.*, 44, W03428, doi:10.1029/2008WR006803.
- Zyvoloski, G.A., Robinson, B.A., Dash, Z.V. and Trease, L.L. (2011) Summary of the Models and Methods for the (FEHM) Application --- A Finite-Element Heat- and Mass-Transfer Code. LA-13306-MS, Los Alamos National Laboratory, Los Alamos, New Mexico.

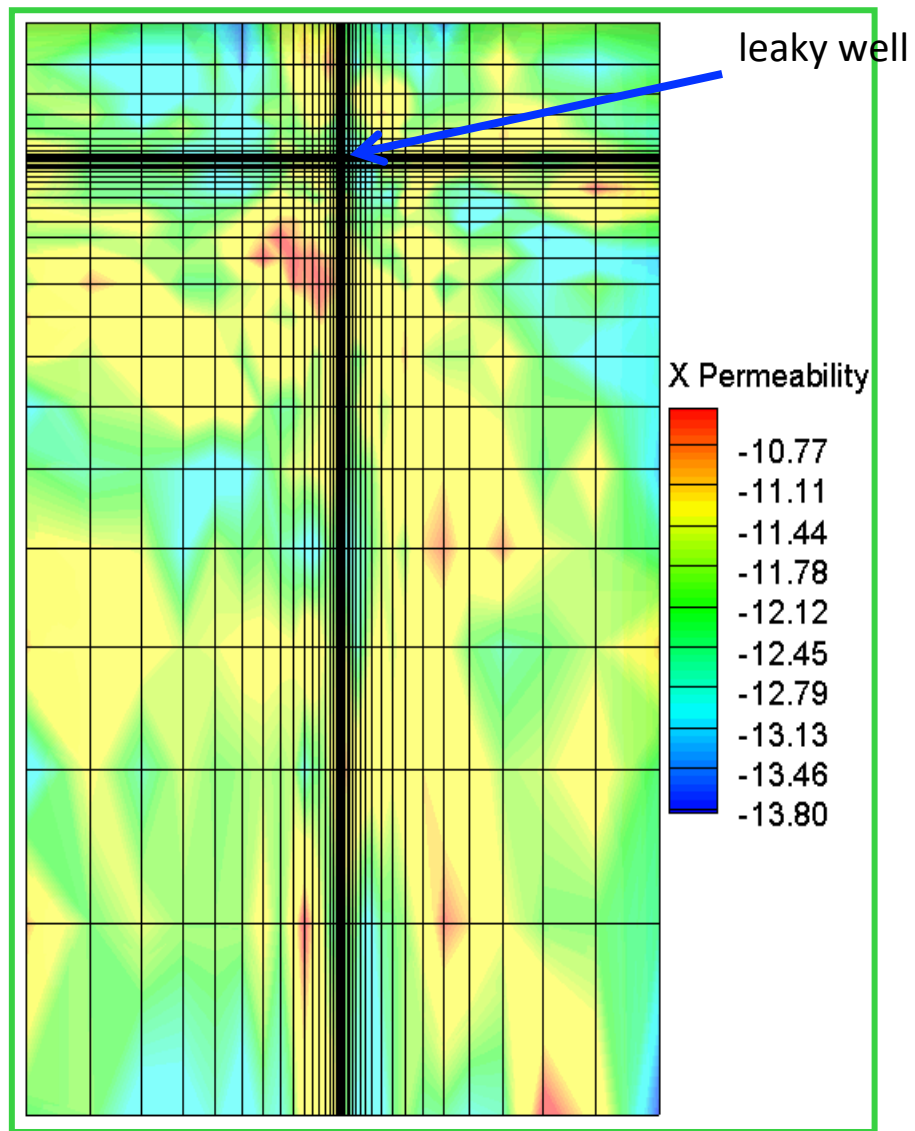


Figure 1. Plan view of the computational mesh, and one realization of permeability heterogeneity.

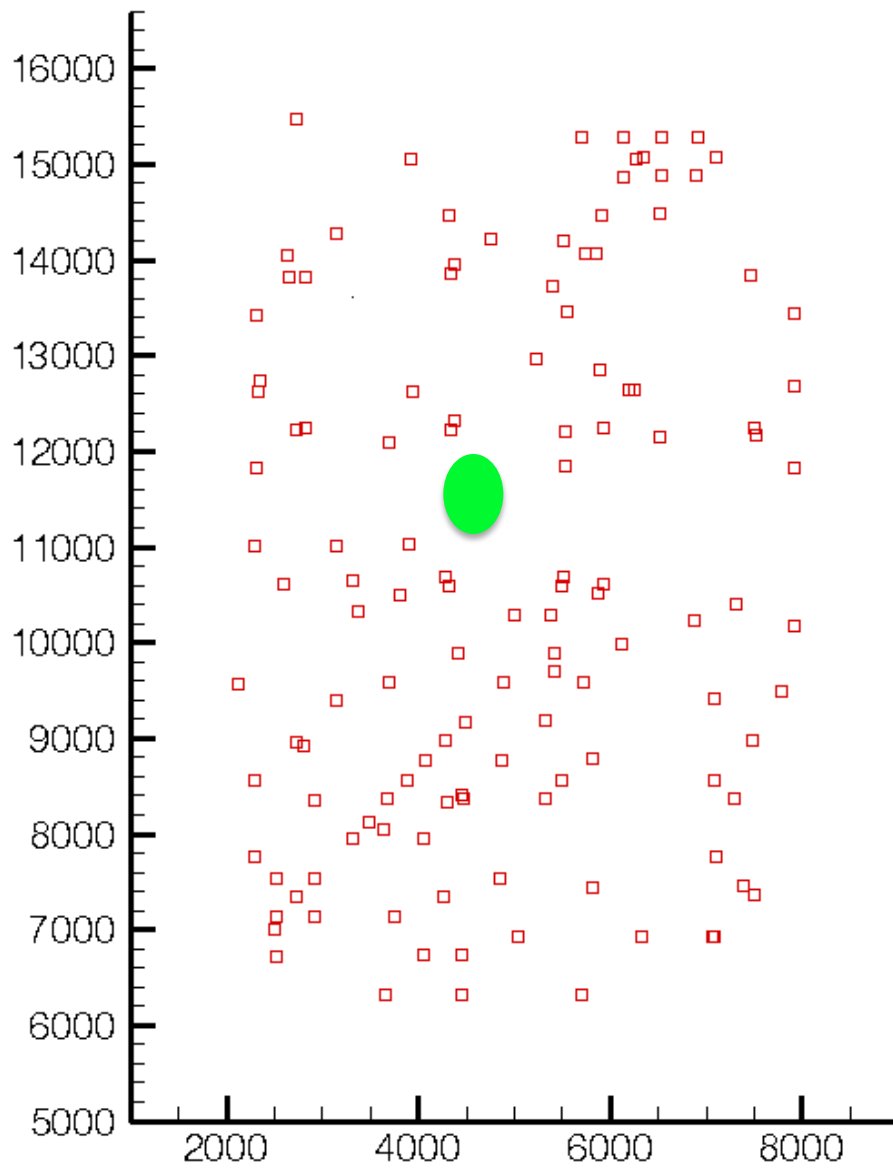


Figure 2. Location of potential leaky wells. For comparison, green ellipse shows approximate size of maximum TDS plume calculated from our simulations.

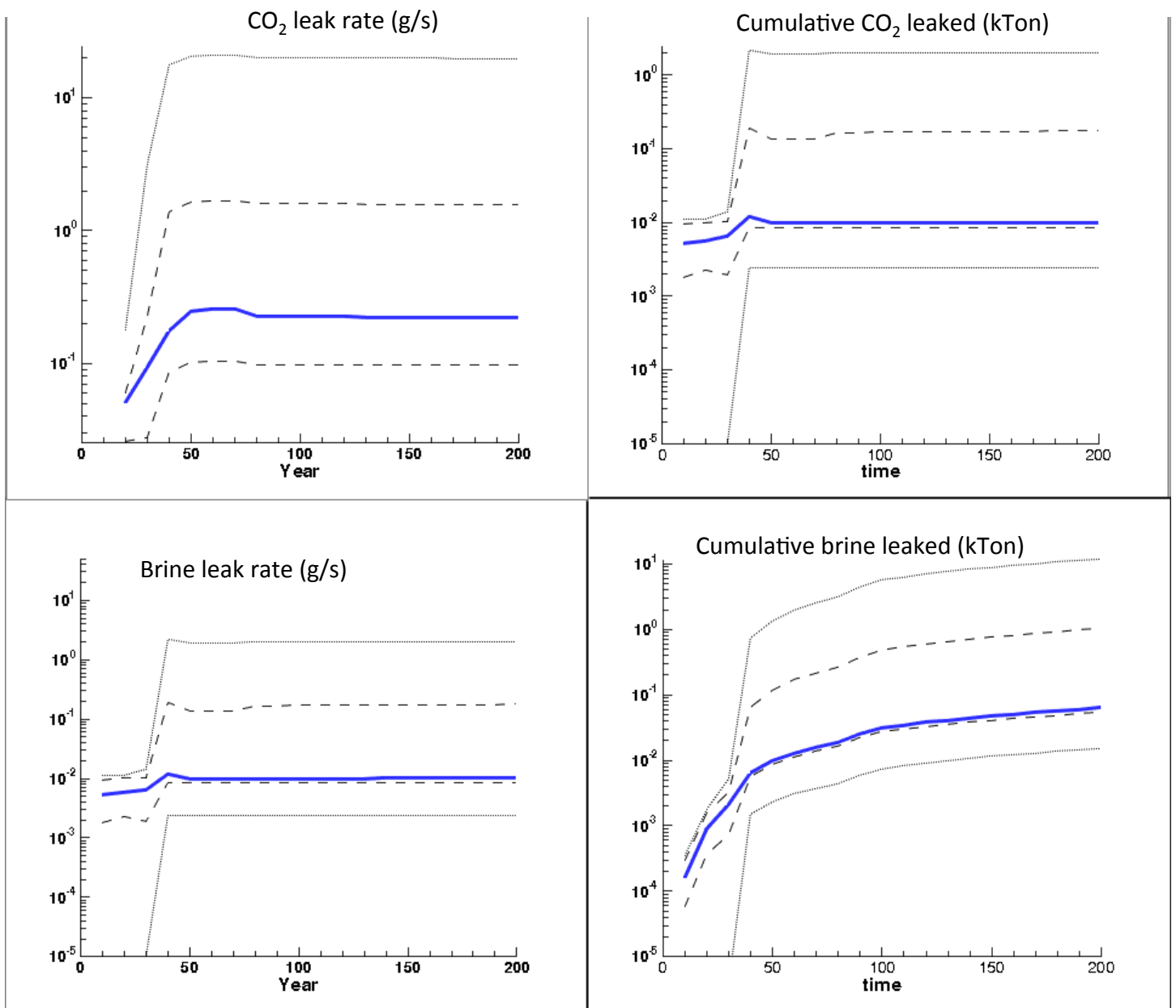


Figure 3. Range of leak rates used in simulations. Blue line: median, dashed line, 25, 75%, grey line: 9% and 95%.

New figure inserted March 19

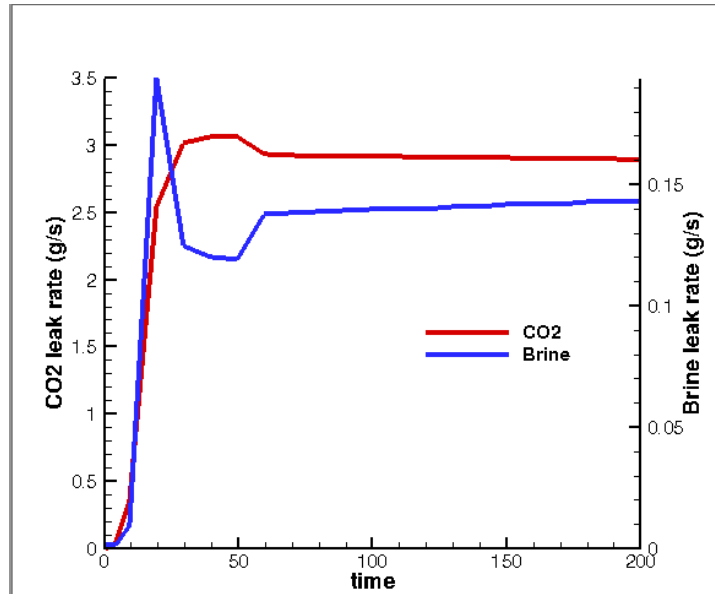


Figure 4. Brine and CO<sub>2</sub> leak rates for example simulation (case 599)

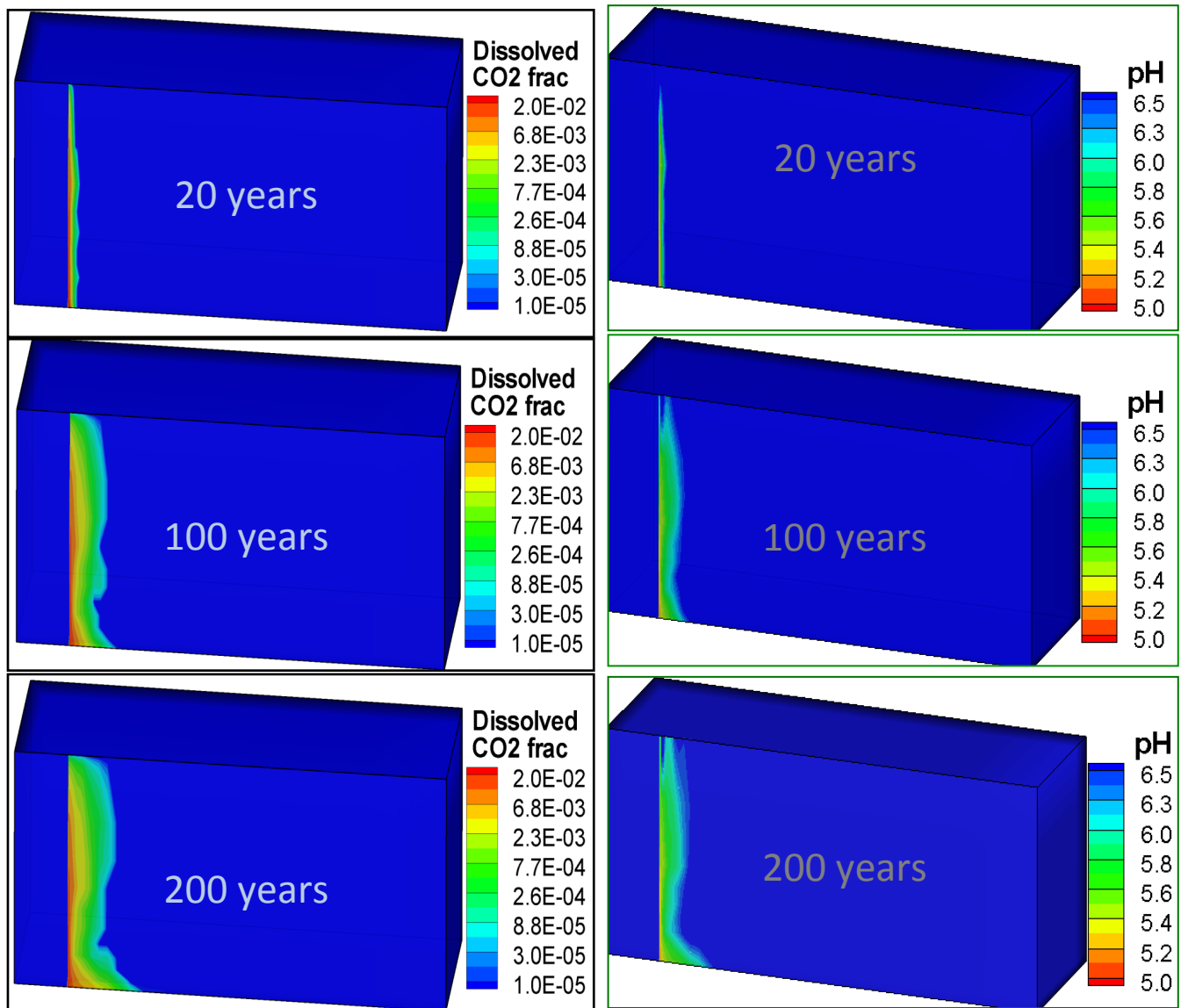


Figure 5. Time snapshots of plumes for example problem

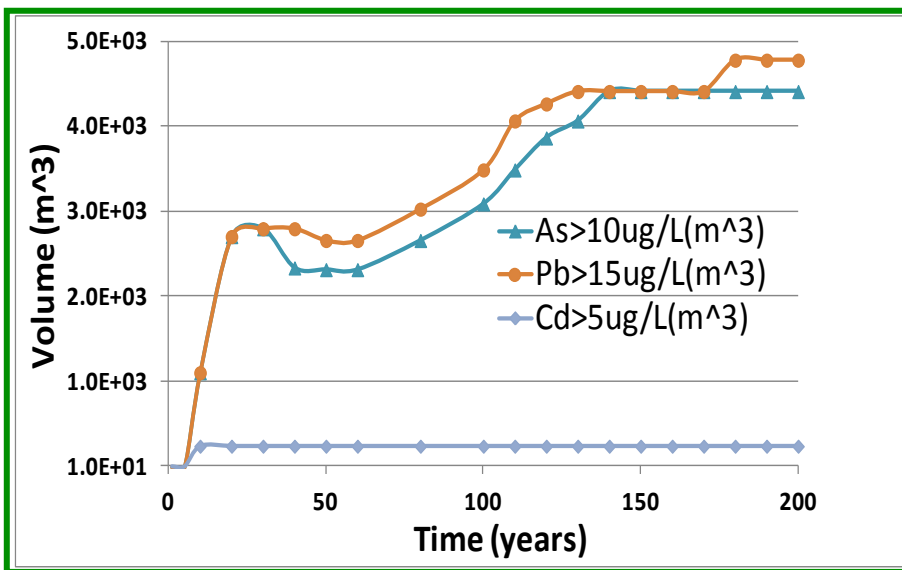
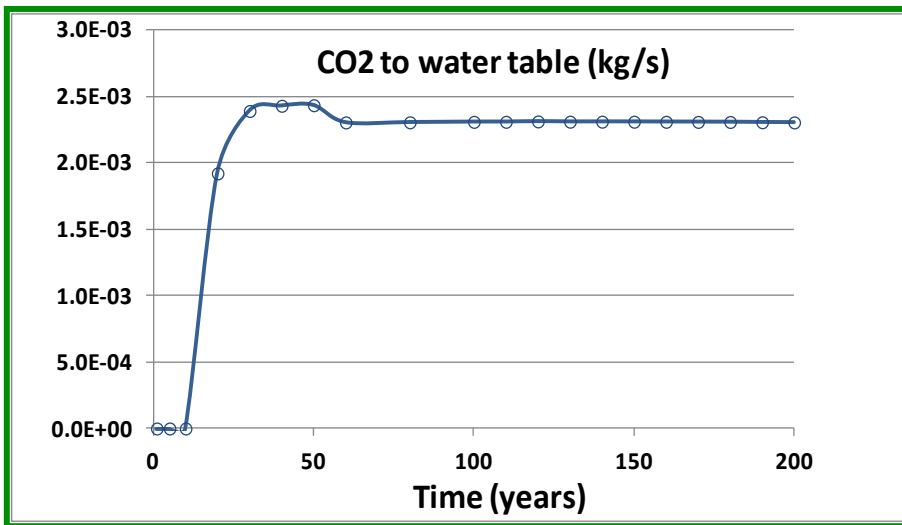
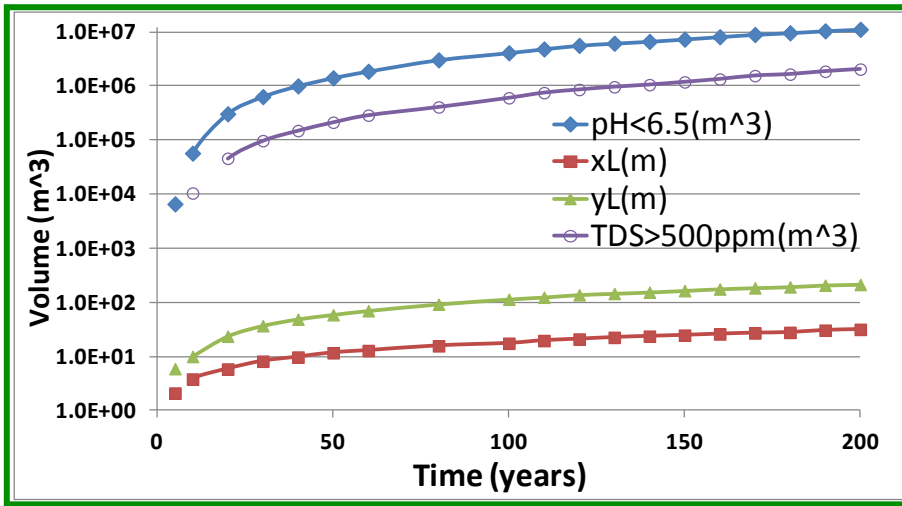


Figure 6. Time varying metrics for case 599

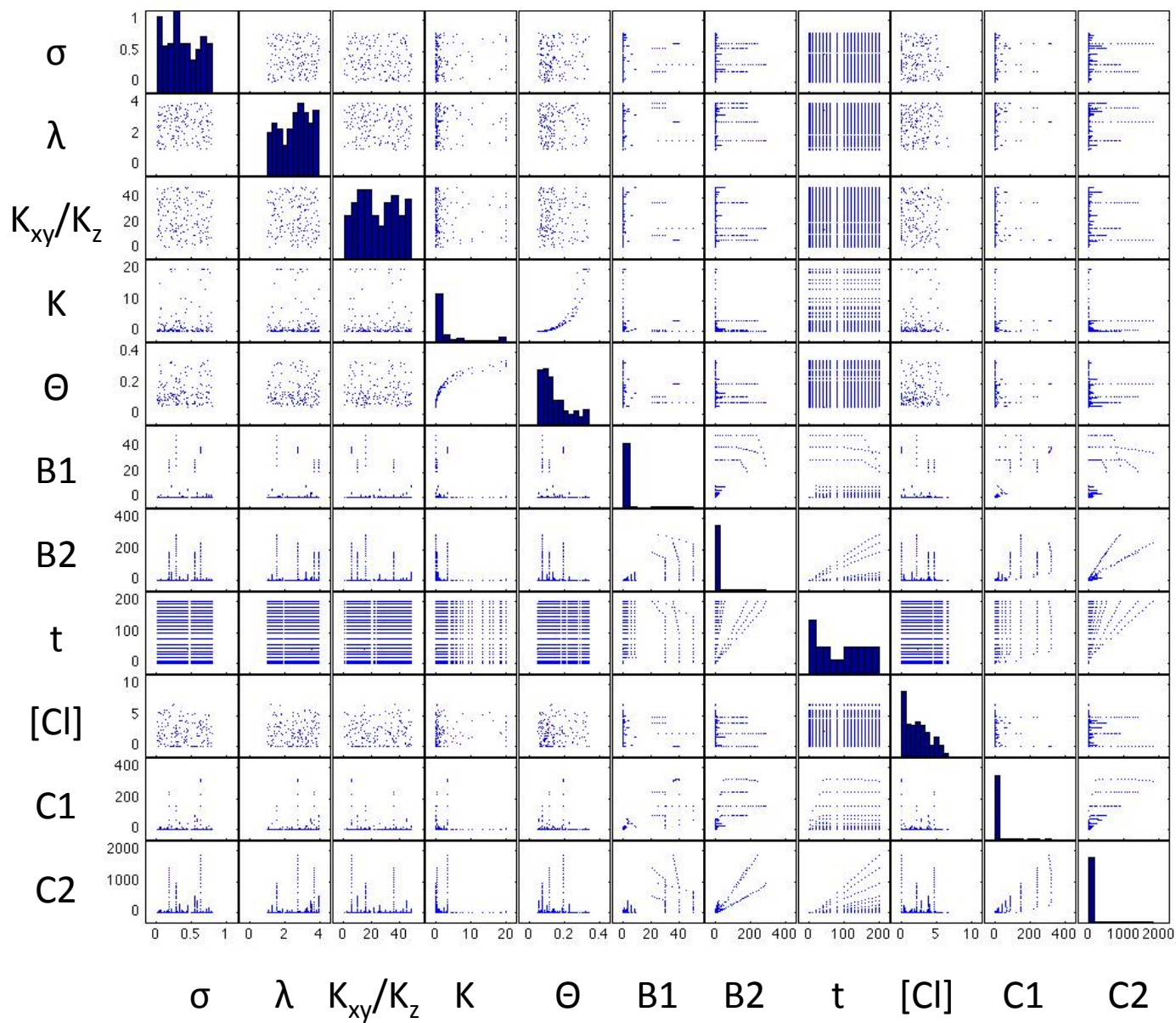


Figure 7. Input parameters used for Monte Carlo runs



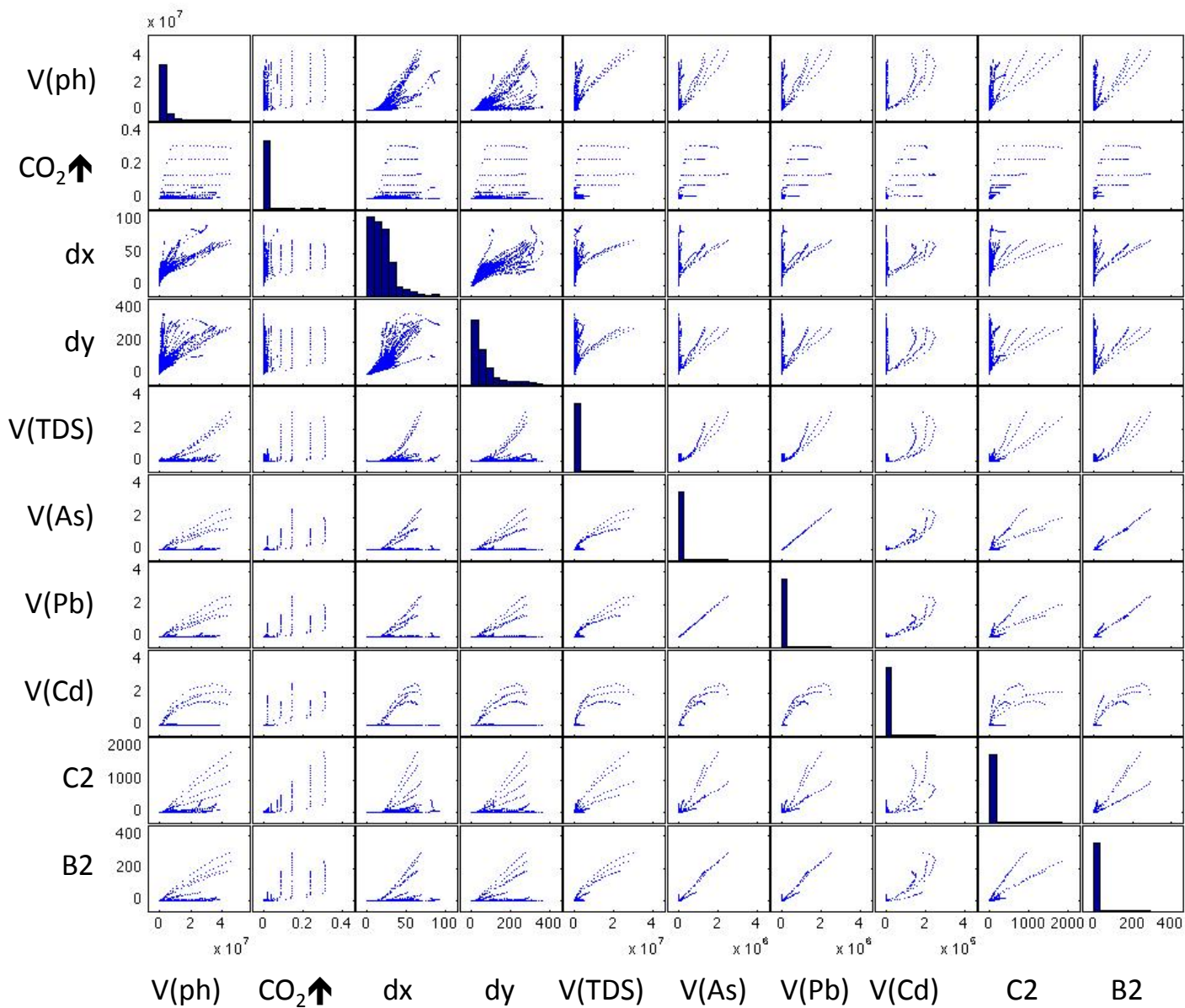


Figure 8a. Model outputs from Monte Carlo runs, using  $C^* = \text{MCL}$  values.

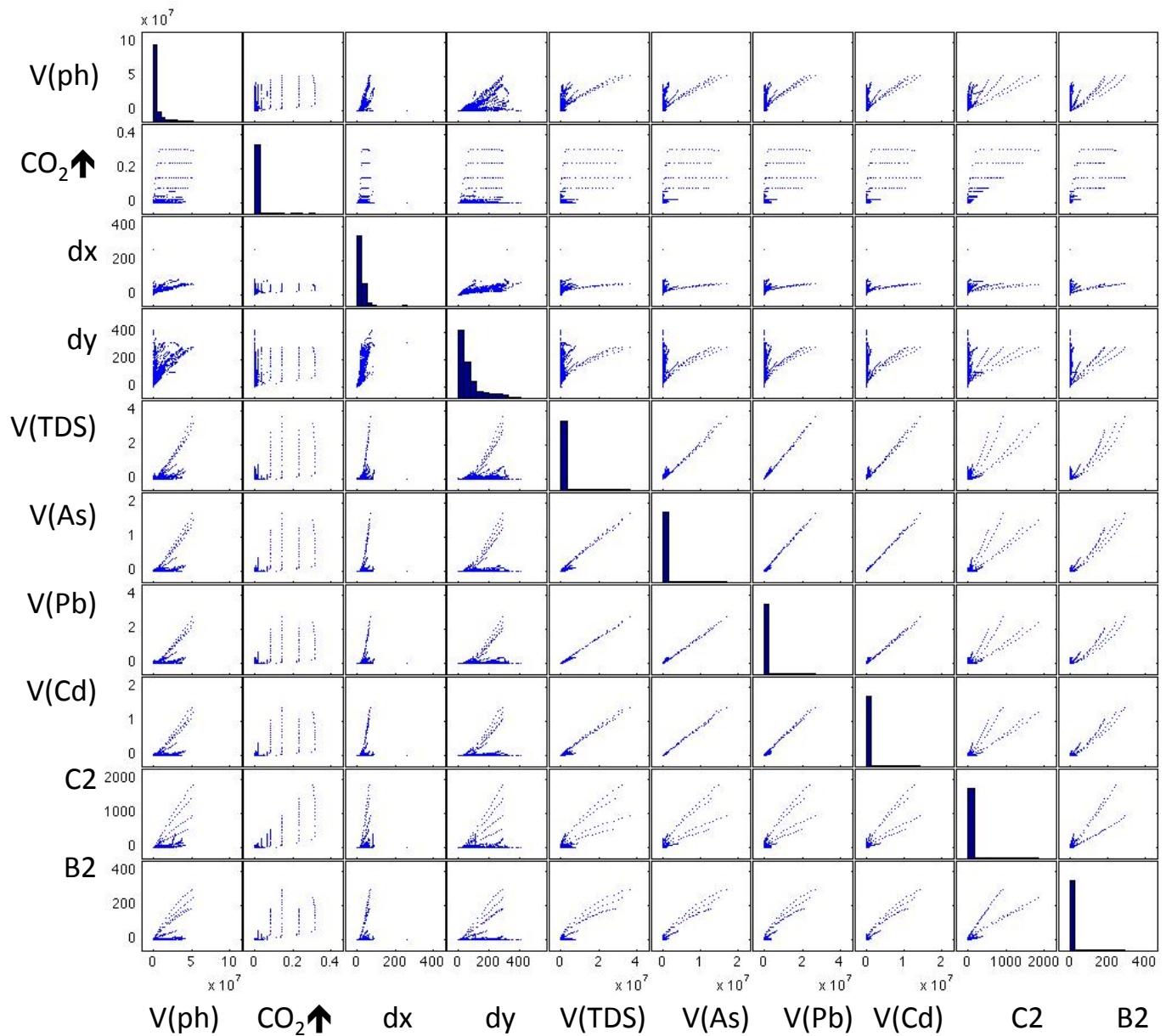


Figure 8b. Model outputs from Monte Carlo runs, using  $C^* = \text{background values}$ .

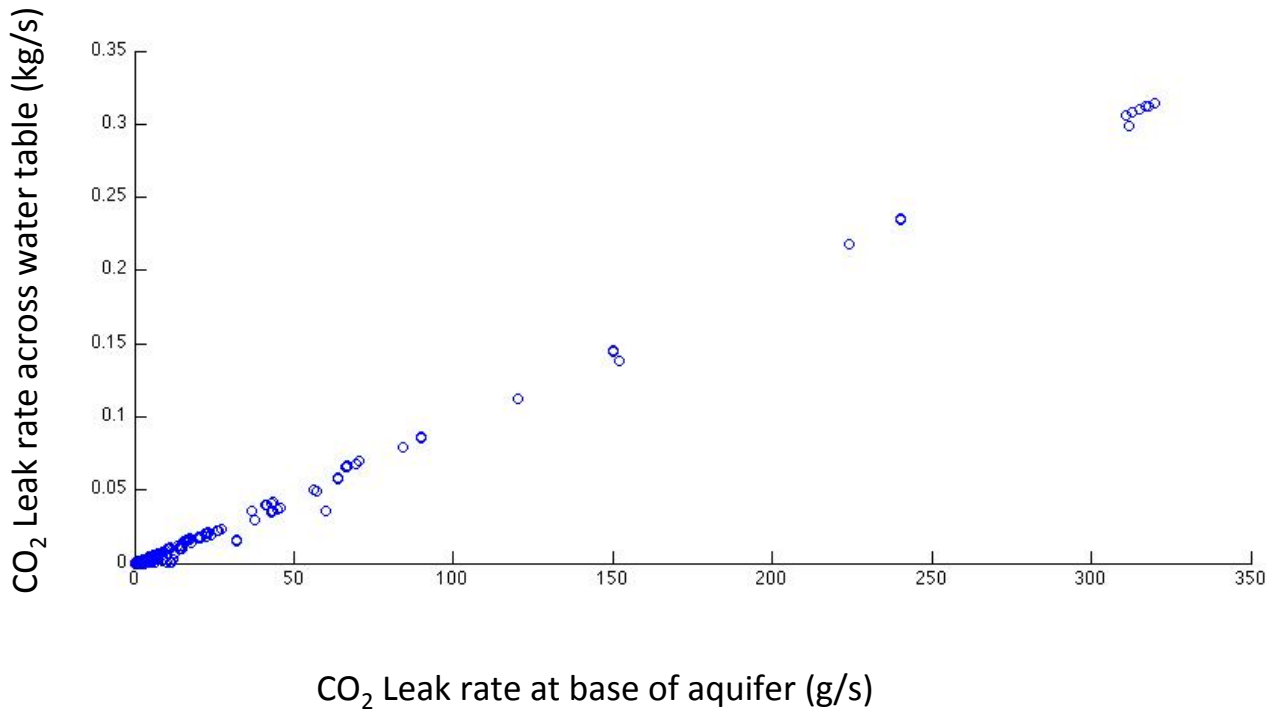


Figure 9. Relationship between CO<sub>2</sub> leak rate (input parameter) and CO<sub>2</sub> flux across the water table (output parameter).

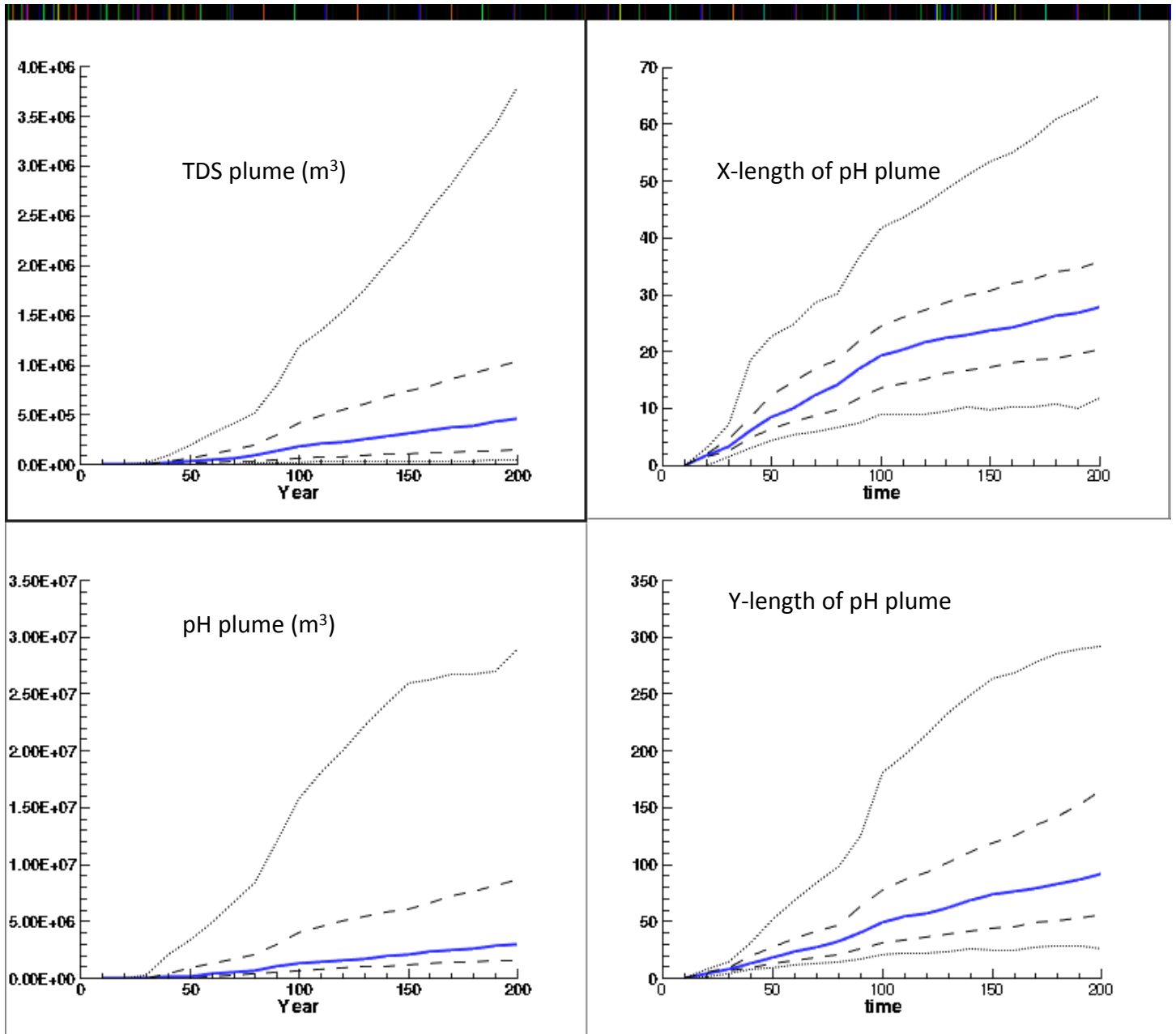


Figure 10a. Time varying simulation outputs (TDS and pH plume metrics) for MCL thresholds. Blue line: median, dashed line, 25, 75%, grey line: 9% and 95%.

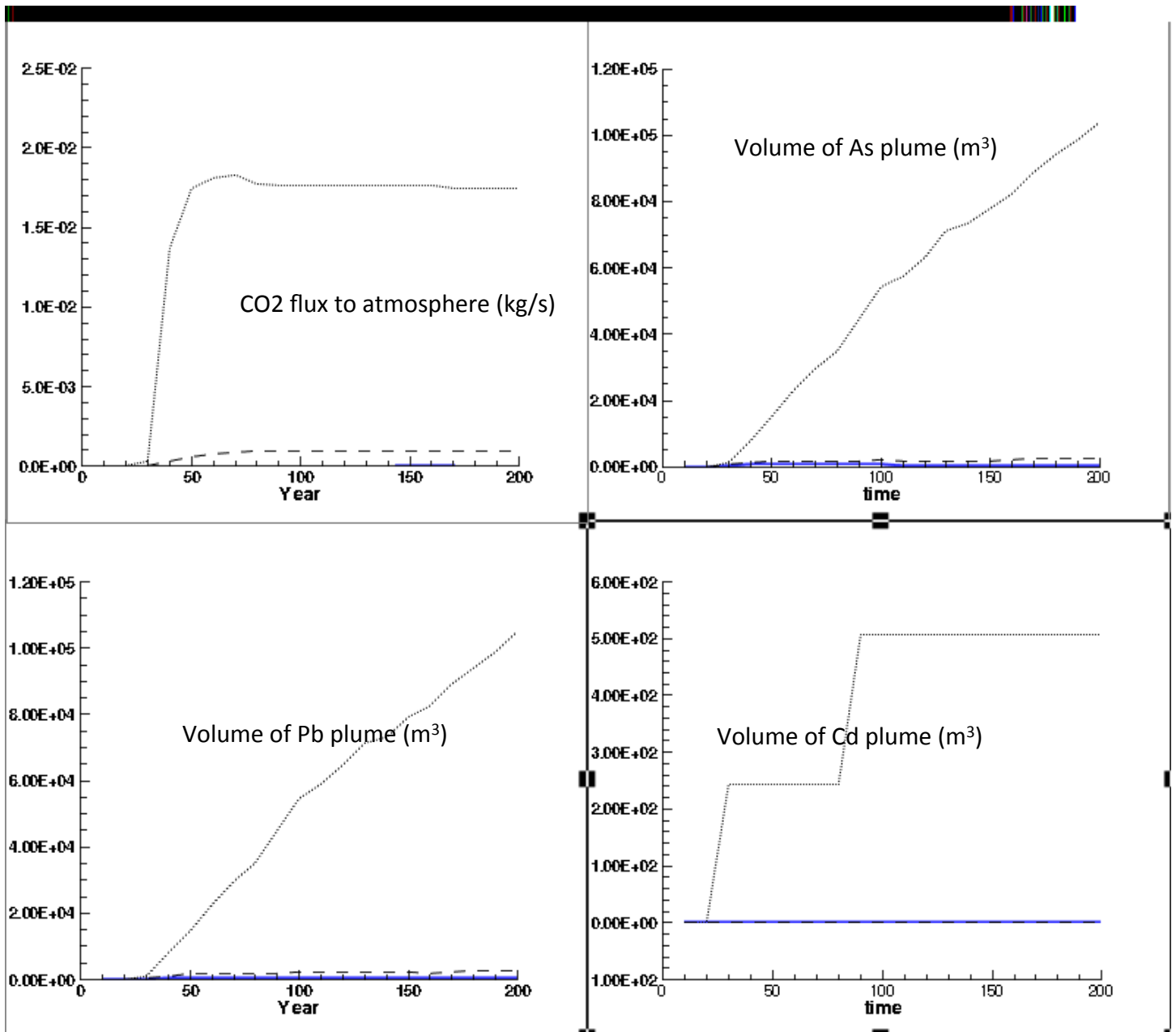


Figure 10b. Time varying simulation outputs (CO<sub>2</sub> flux and trace metal plumes) for MCL thresholds. Blue line: median, dashed line, 25, 75%, grey line: 9% and 95%.

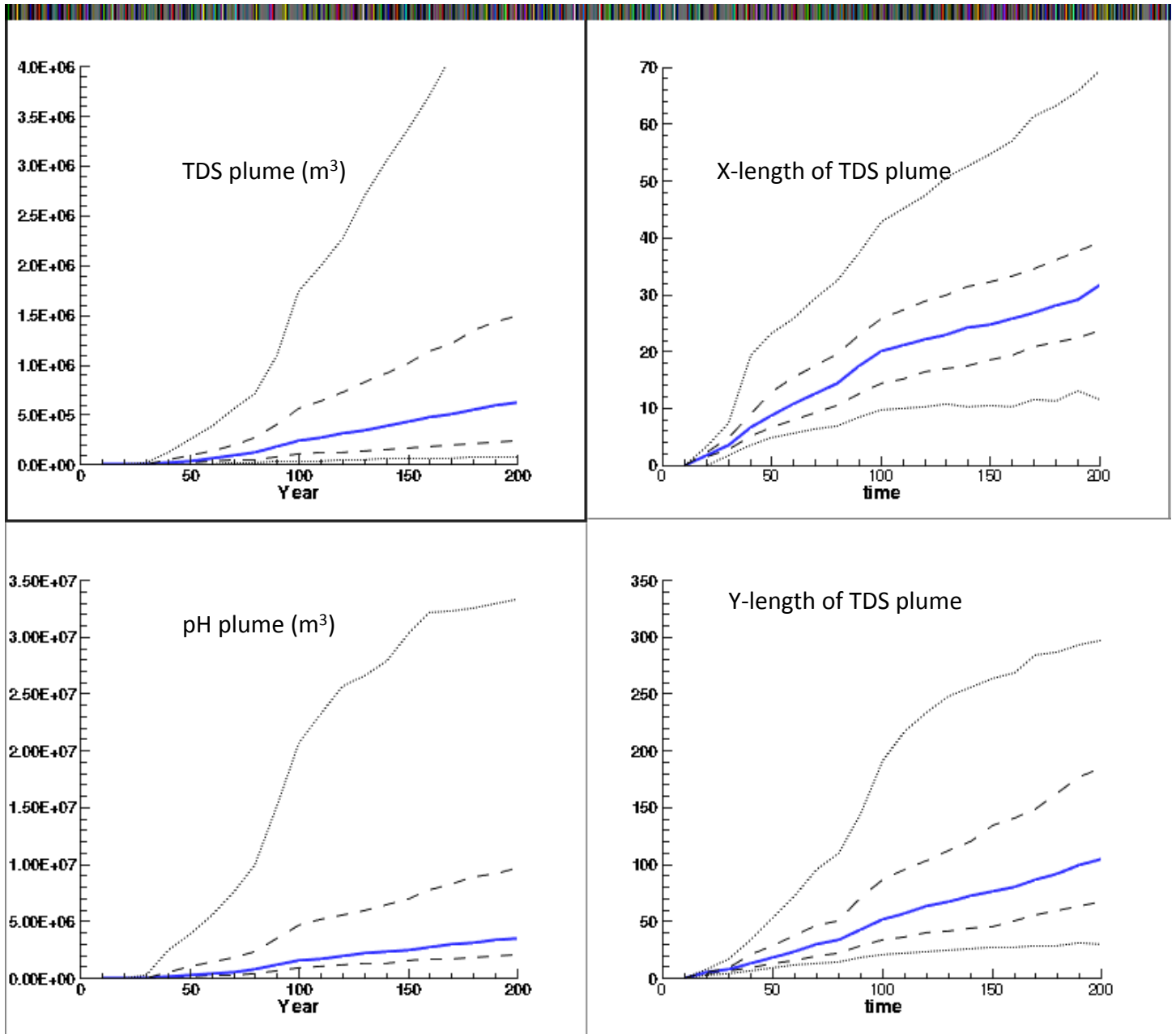


Figure 10c. Time varying simulation outputs (TDS and pH plume metrics) for background critical threshold values. Blue line: median, dashed line, 25, 75%, grey line: 9% and 95%.

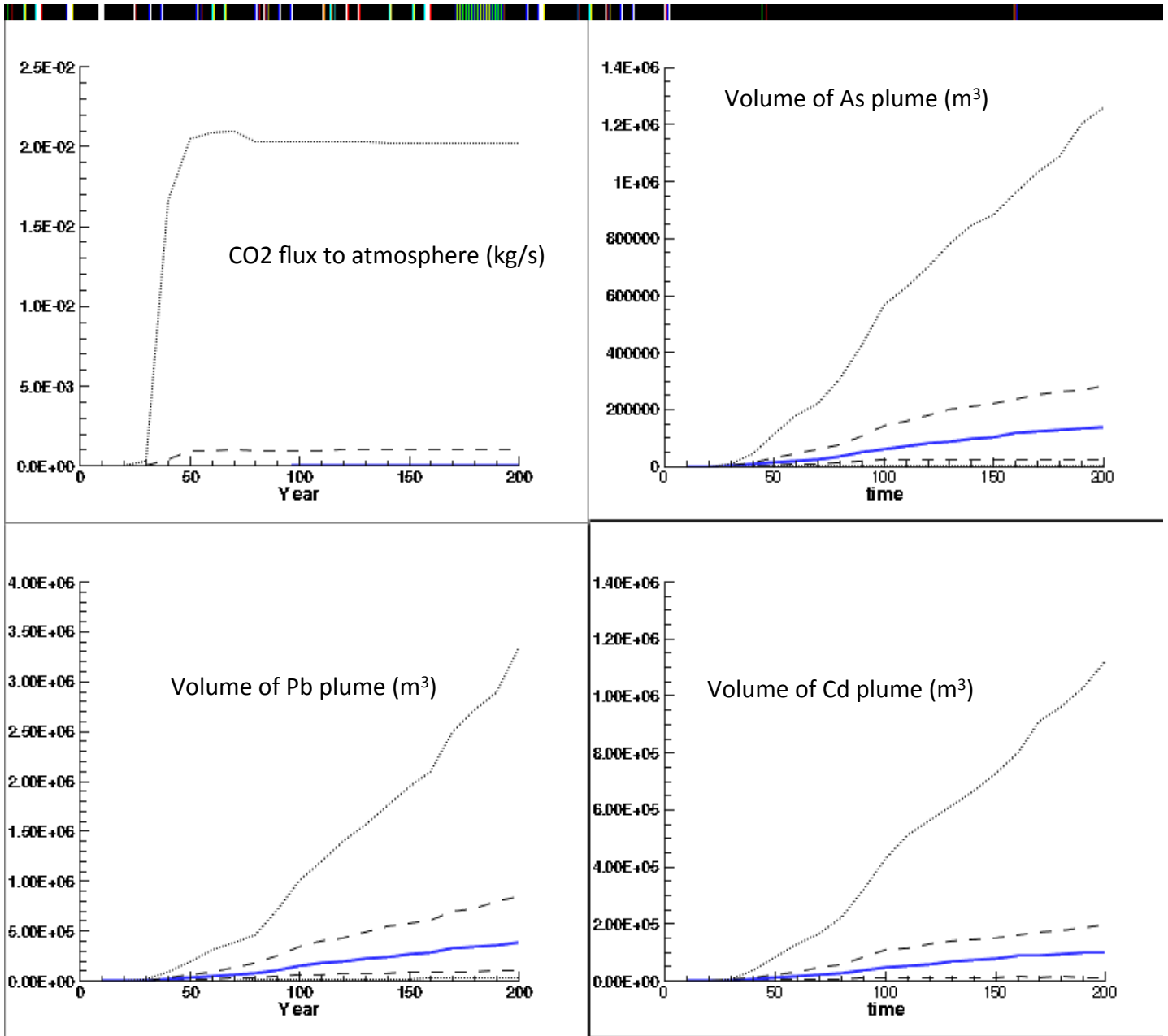


Figure 10d. Time varying simulation outputs (CO<sub>2</sub> flux and trace metal plumes) for background threshold values. Blue line: median, dashed line, 25, 75%, grey line: 9% and 95%.

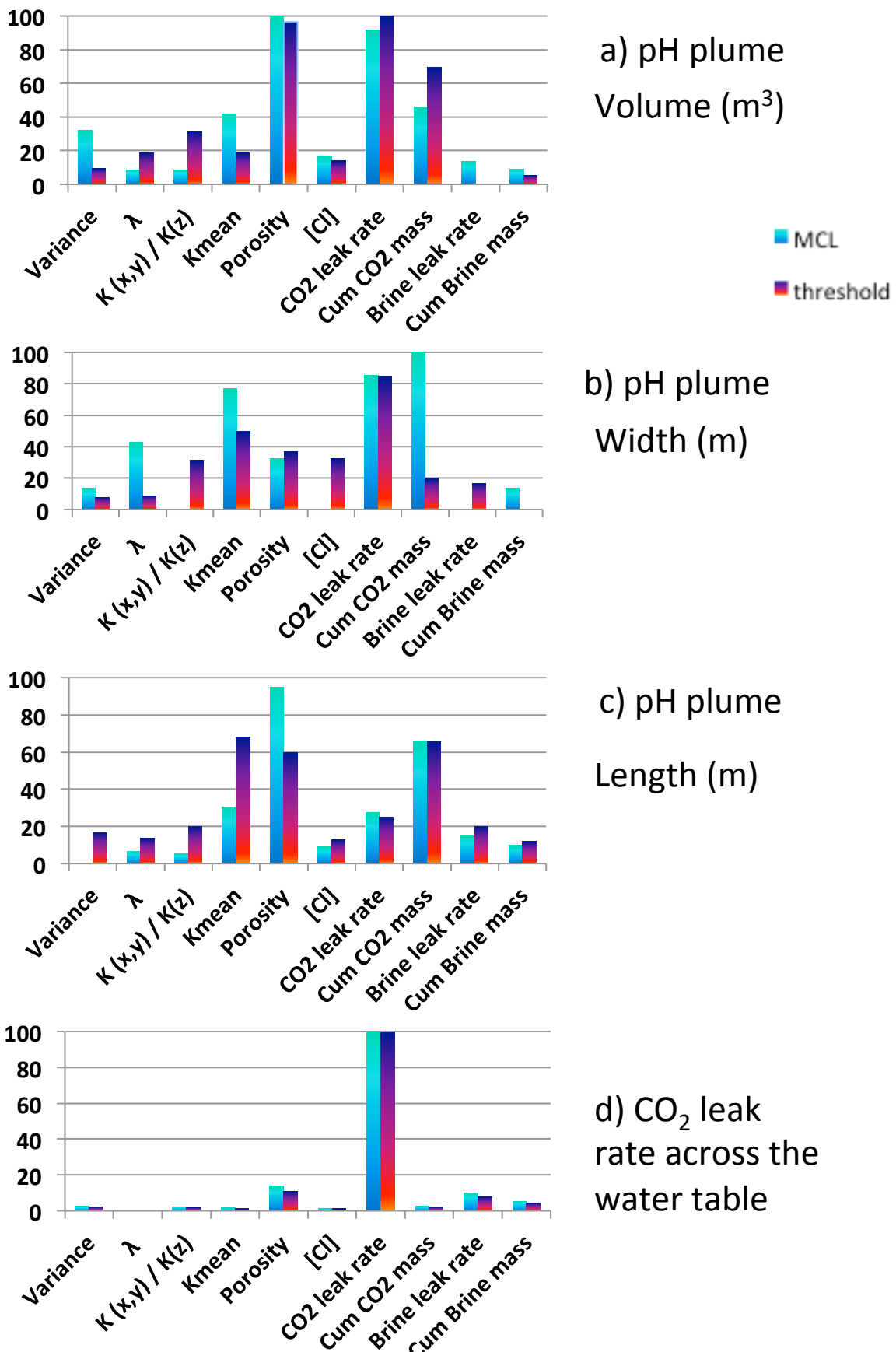


Figure 11. Importance measure for 11 model input parameters and 4 model outputs: pH plume characteristics (a-c) and CO<sub>2</sub> leak rate across the water table.



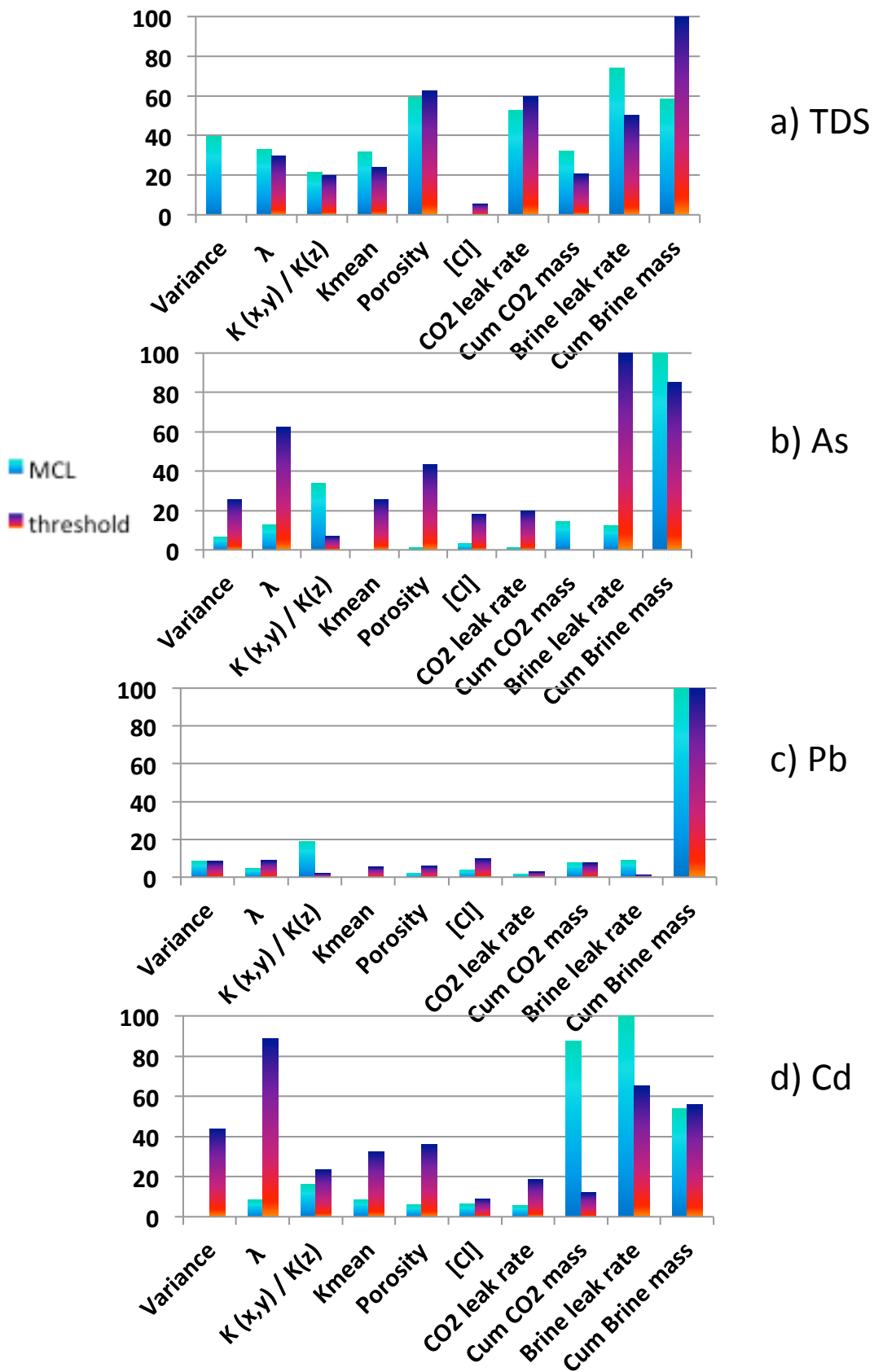


Figure 12. Importance measure for 11 model input parameters and 4 model outputs: TDS, As, Pb and Cd plume volumes.

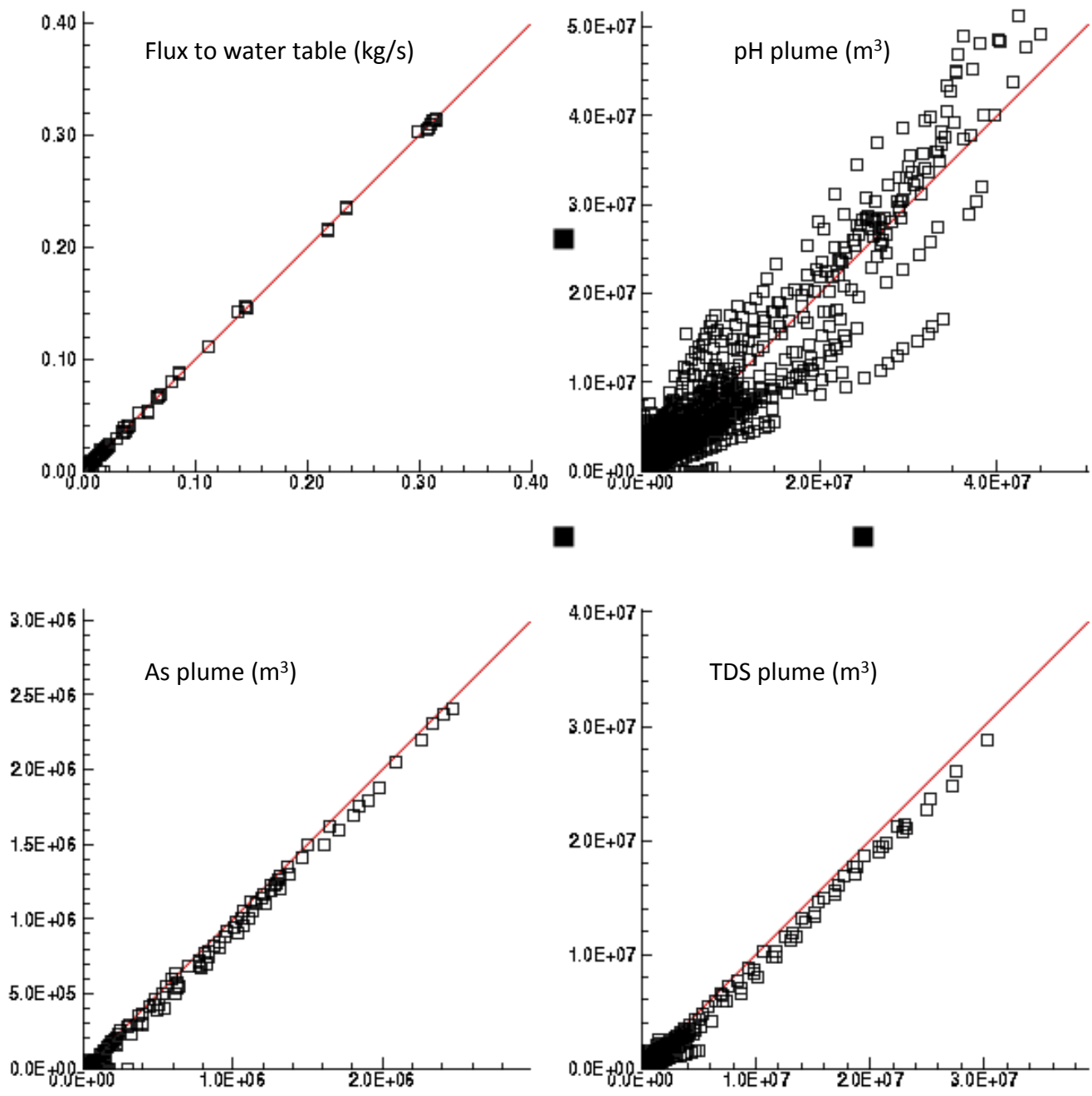


Figure 13a. Comparison of simulations and from the quadratic response surfaces for four key metrics, using the MCL standards. In all cases the horizontal axis is the process model results, the vertical axis is the response surface prediction. The red line indicates a perfect fit.

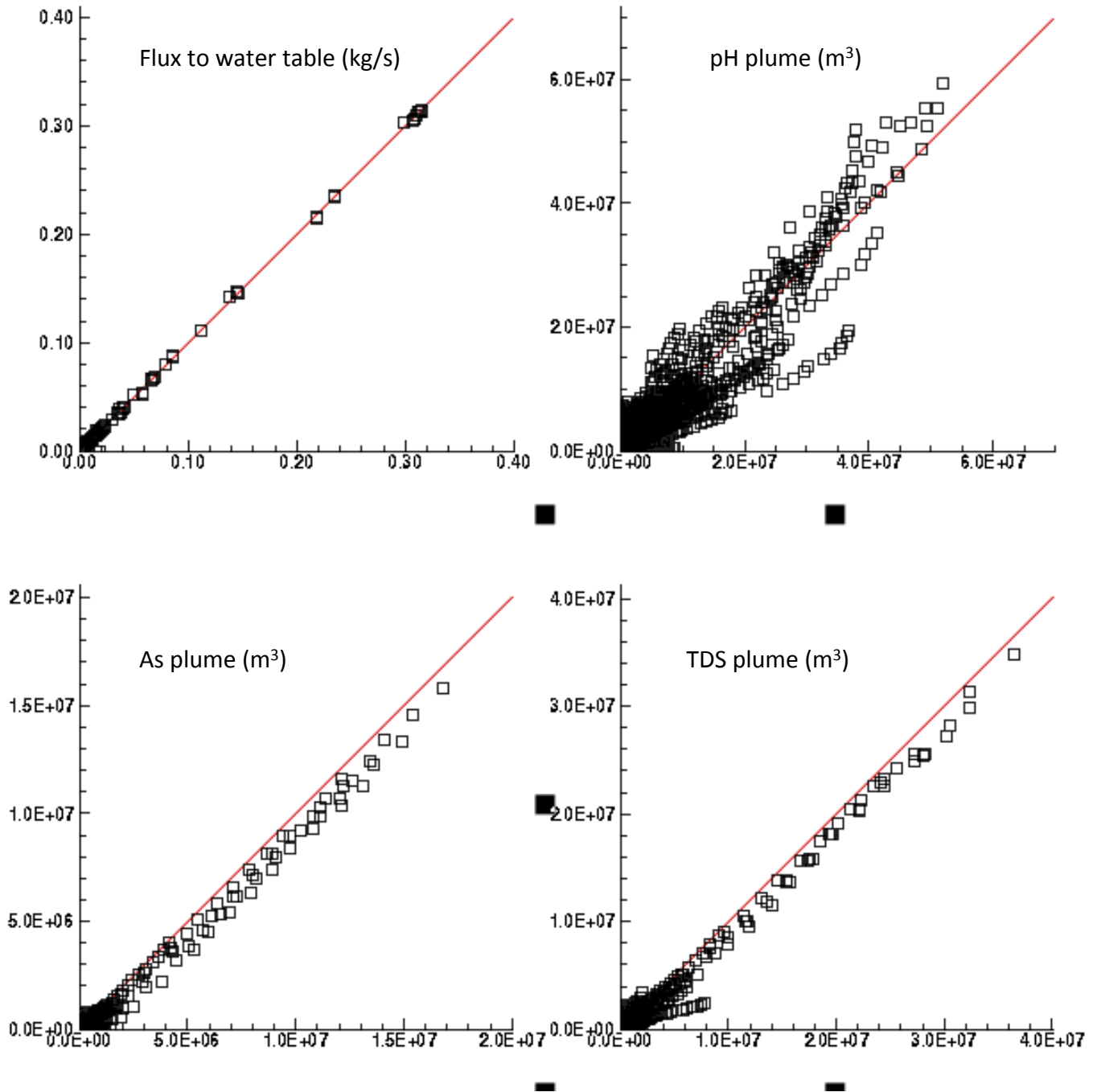


Figure 13b. Comparison of simulations and from the cubic response surfaces for four key metrics, using the threshold standards. In all cases the horizontal axis is the process model results, the vertical axis is the response surface prediction. The red line indicates a perfect fit.

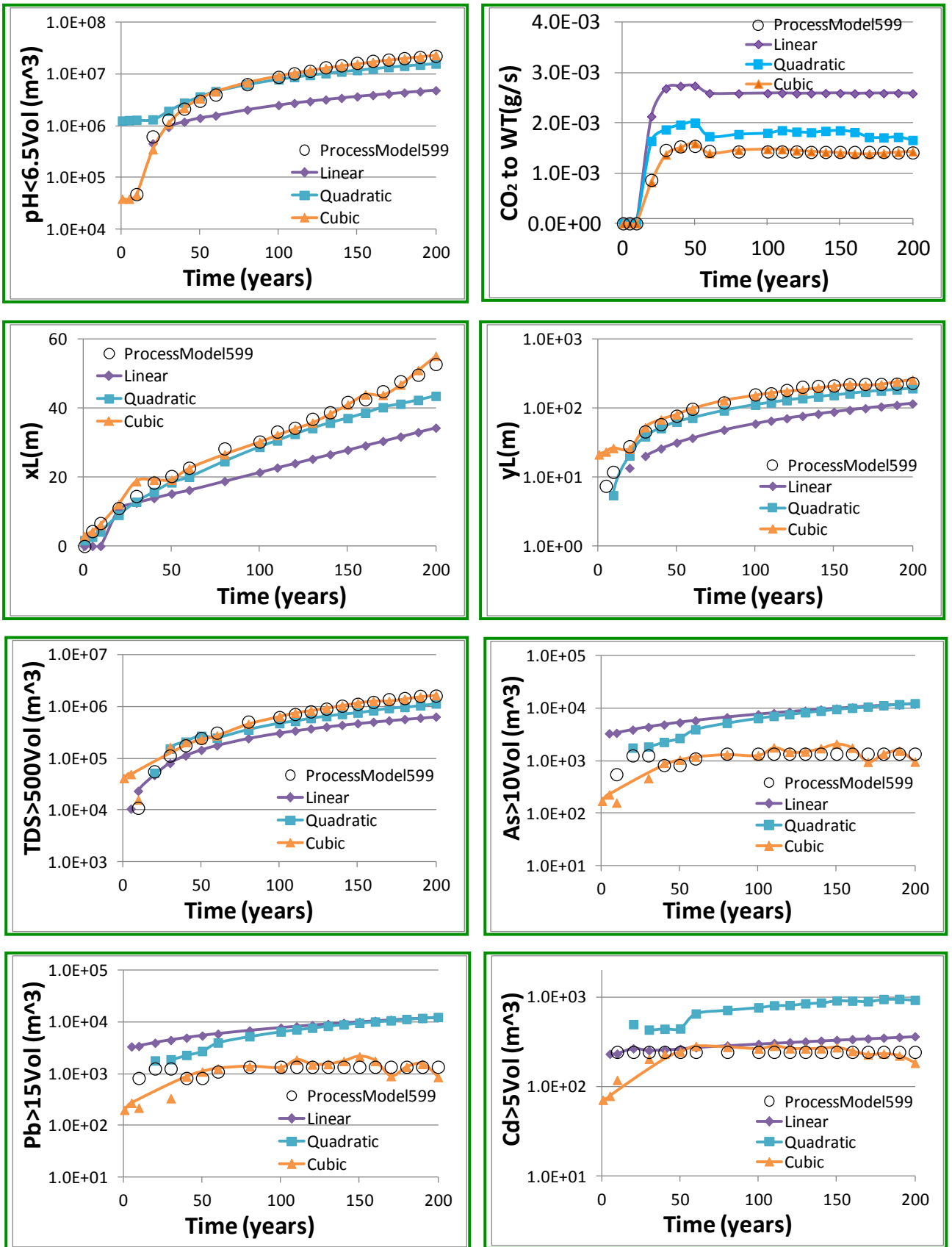




Figure 14. Comparison time-varying metrics for process model and response surface outputs for case 599 (shown in Figure 5).

## Systems model

- 
- 1- 5 aquifer parameters
  - 2- [Cl] of leaky brine
  - 3- Number of leaking wells (N)
  - 4- time
  - 5- For each well,
    - a- XY location
    - b- CO<sub>2</sub> Leakage rate (kg/s)
    - c- Cumulative CO<sub>2</sub> leakage (kg)
    - d- Brine leakage rate (kg/s)
    - e- Cumulative brine leakage (kg)

- 
- 1- Total flux to wt (kg/s)
  - 2- Volume of plumes (5) (m<sup>3</sup>)
  - 3- X,Y dimension of TDS plumes

## GW ROM

1

Loop over N wells, If any pair of wells are too close <sup>1</sup>, combine into a single leaky well and sum leakage rates. Result: list of M wells and leakage rates, where  $M \leq N$

2

Loop over M leaks

Has leak arrived at well? If yes:

Calculate  $t_1$ =time since CO<sub>2</sub> leak began,  $t_2$ =time since brine leak began  
Call response surface

12 Inputs: leakage rate (brine and CO<sub>2</sub>), cumulative leak mass (brine and CO<sub>2</sub>), 5 aquifer parameters, [Cl] of brine,  $t_1, t_2$

8 Outputs: total flux to wt, volume of 5 plumes, xy dimension of TDS plume

Optional: check to ensure xy dimension of TDS plume meets superposition criterion, if not, return to step 1 with refined Rmin

Calculate running sum of total flux and volume

<sup>1</sup> $R < R_{min}$ , where  $R_{min}$  is defined *a priori* by examination of all possible ROM results

Figure 11. Relationship between GW Reduced order model and systems model