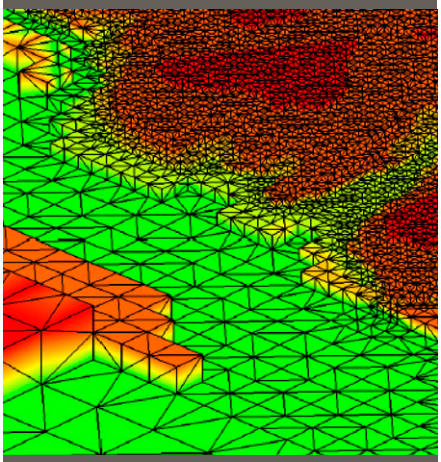
Special Section: Contaminants in the Vadose Zone

Philip H. Stauffer*
Zhiming Lu



A new unsaturated Monte Carlo sampling strategy, the retention curve subsampling algorithm, can reduce the number of Latin hypercube samples required for convergence to the mean and 5 and 95% confidence bounds of particle breakthrough by sampling across one-dimensional retention curve space instead of blindly combining five dimensional van Genuchten parameters.

Computational Earth Sciences Group (EES-16), Los Alamos National Lab., Los Alamos, NM 87545. *Corresponding author (stauffer@lanl.gov).

Vadose Zone J.
doi:10.2136/vzj2011.0171
Received 16 Nov. 2011.

© Soil Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA.
All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Quantifying Transport Uncertainty in Unsaturated Rock using Monte Carlo Sampling of Retention Curves

We have developed a new Monte Carlo sampling method for simulating flow and transport in unsaturated porous media, characterized by van Genuchten–Mualem constitutive relations. Instead of sampling each individual soil parameter from its probability space and then running Monte Carlo simulations using realizations of rock parameters directly, we calculate retention curves from realizations of rock parameters, take subsamples from these retention curves, and run simulations using parameter realizations corresponding to these selected retention curves. The retention curve subsampling methodology was applied to three-dimensional simulations of conservative tracer transport beneath Material Disposal Area G at the Los Alamos National Laboratory. Convergence of the proposed sampling method was assessed by comparing statistics of breakthrough curves observed at a compliance boundary with those obtained using between 25 and 1000 Latin hypercube sampling (LHS) Monte Carlo simulations. Our example shows that 25 model runs based on selected retention curves could adequately approximate the results from our assumed truth (1000 LHS Monte Carlo simulations), while LHS alone required in excess of 50 realizations to achieve the same quality result. Another finding from this work is that the median of the breakthrough curves was more meaningful than the arithmetic mean of curves, and the former was nearly identical to the breakthrough curve derived from mean rock properties.

Abbreviations: BTC, breakthrough curve; FEHM, finite-element heat- and mass-transfer code; LANL, Los Alamos National Laboratory; LHS, Latin hypercube sampling; MDA, Material Disposal Area; RCS, retention curve subsampling.

Simulated breakthrough curves (BTCs) at compliance boundaries are often used to predict transport behavior at contaminated sites (Painter et al., 2001). In cases that involve uncertainties in input parameters, such as water flux at boundaries or rock properties, Monte Carlo simulation (MCS) is commonly used to explore uncertainty in the resulting BTCs (Riva et al., 2008). In MCS, a large number of realizations of uncertain parameters are generated, the flow or transport equations are solved for each individual realization, and the statistics of the BTCs are computed from all realizations (Zhang et al., 2007).

Monte Carlo simulation is computationally very demanding, especially on large computational meshes representing unsaturated media where the flow equation is nonlinear. We hypothesized that, to reduce the complexity of unsaturated MCS, we could sample from a distribution of retention curves derived from many realizations of rock parameters (saturated permeability and van Genuchten fitting parameters) instead of sampling individual parameters that go into the curves. In other words, instead of sampling each individual rock parameter from its probability space and then running Monte Carlo simulations using realizations of rock parameters directly, retention curves may be calculated from realizations of rock parameters, subsamples taken from these retention curves, and simulations run using parameter realizations corresponding to these selected retention curves. The elegance of the retention curve subsampling (RSC) approach is that it greatly reduces the dimensions of uncertainty in the problem and results in far fewer simulations needed to span the behavior of the system while retaining the same information as the more complex and time-consuming method of sampling all uncertain parameters randomly.

We investigated the RCS technique using three-dimensional simulations of conservative tracer BTCs related to an ongoing site analysis at Material Disposal Area (MDA) G, Los Alamos National Laboratory (LANL). One thousand full Monte Carlo simulations were used as "truth" and convergence of the proposed method toward this "truth" was compared

with convergence of a traditional Latin hypercube sampling method.

Site Description

Material Disposal Area G, the only active low-level waste repository for LANL, has been in operation since 1957. Material Disposal Area G is located approximately 2 km west of the town of White Rock, NM, and about 5 km west of the Rio Grande. The site lies on Mesita del Buey, a finger mesa on the Pajarito Plateau along the eastern flank of the Valles Caldera. The surface of MDA G slopes from west to east from an elevation of 2070 m asl to an elevation of approximately 2033 m asl (Fig. 1).

Subsurface information about the basic stratigraphy beneath MDA G, obtained from a variety of both shallow site wells (<70 m) and regional characterization wells (>400 m), is shown in cross-section in Fig. 2. The hydrogeology of the Pajarito Plateau was described by Broxton and Vaniman (2005), with a more detailed description of the unsaturated geologic units by Stauffer et al. (2005a) and Vrugt et al. (2008). The waste disposal pits and shafts at MDA G have been excavated into Unit Qbt2 and Unit Qbt1 of the Tshirege Member of the Bandelier Tuff, which extends below the ground surface to approximately 1970 m asl in the vicinity of MDA G. Beneath this lies Unit Qbt1 g and the basal unit of the Tshirege Member, the Tsankawi Pumice (Qtt), while the Cerro Toledo interval (Qct), a unit composed of poorly sorted volcanic sediment, lies between the Tshirege Member and the upper Otowi Member (Qbof) of the Bandelier Tuff. The basal unit of the Otowi Member is the Guaje Pumice (Qbog). The total thickness of the Bandelier Tuff generally increases from east to west in the vicinity of MDA G; however, large variation in the thickness of individual units occurs with some units, such as Qtt, Qct, and Qbog, which are entirely absent at certain locations. Beneath these rhyolitic tuffs lies the very thick Cerros del Rio basalt (Tb4) that extends to the water table at a depth of approximately the 300 m

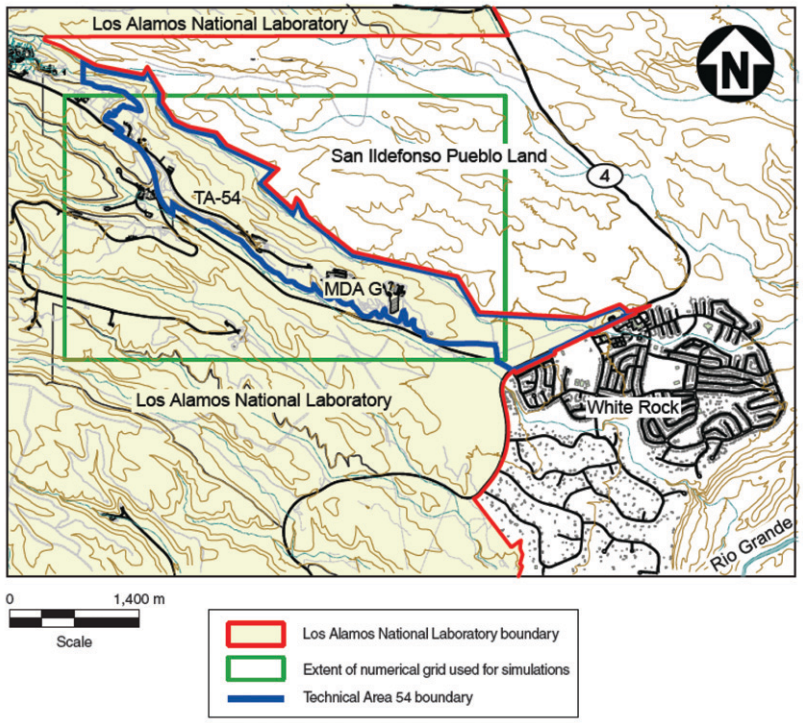


Fig. 1. Location of Material Disposal Area (MDA) G in relationship to the town of White Rock, NM. The green box shows the outline of our model domain (image from Stauffer et al., 2005b).

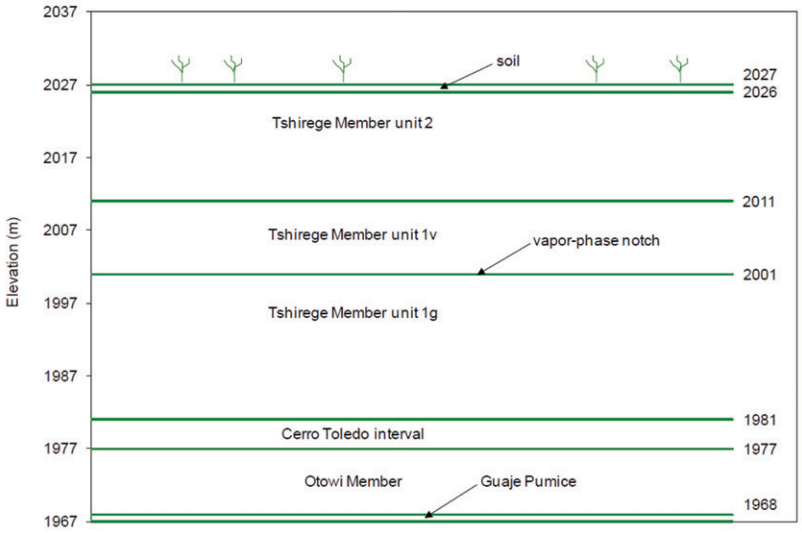


Fig. 2. Simplified cross-section showing the stratigraphic units through Material Disposal Area G from the surface to the bottom of the Bandelier Tuff. Geologic units below this have virtually no impact on particle breakthrough (image from Levitt, 2011).

Numerical Model Description

The hydrogeologic modeling used in this study was adapted from Stauffer et al. (2005b), who investigated the potential migration of contaminants from MDA G toward a compliance boundary lying 100 m to the east of the site. This work was built based on a number of earlier investigations, including performance assessment and composite analysis (PA/CA) (Hollis et al., 1997), a geologic framework (Vaniman et al., 1996; Broxton and Vaniman, 2005), statistical analysis of the hydrologic parameters for MDA G (Krier et al., 1996; Rogers and Gallaher, 1995), a high-resolution digital elevation model (DEM) of the surface topography (Carey and Cole, 2002), and hydrogeologic modeling by Birdsell et al. (2000) in support of the MDA G PA/CA. The geologic framework in the current analysis was updated in 2009 to include new borehole data collected from 2004 to 2009.

Porous Simulator

All simulations presented used the finite-element heat- and mass-transfer code (FEHM) developed by Zyvoloski et al. (1997, 2007). This multiphase porous simulator has been used extensively for studies in the unsaturated zone (Viswanathan et al., 1998; Stauffer et al., 2005a, 2009; Stauffer and Stone, 2005; Vrugt et al., 2008). Details of the governing equations for multiphase flow were presented by Stauffer and Rosenberg (2000). FEHM is regularly benchmarked against an extensive set of test problems to ensure that continued code evolution does not adversely impact known solutions (Dash, 2003), and a list of >100 publications using FEHM can be found at http://fehm.lanl.gov/pdfs/FEHM_references_list.pdf.

Computational Mesh

The three-dimensional hydrogeologic modeling requires a computational mesh that represents the topography and geology of MDA G and the surrounding area. The computational mesh used to conduct the hydrogeologic modeling was designed to meet several conditions, including continuity and correlation with the Española basin site-scale regional aquifer computational model (Keating et al., 2003), adequately high resolution to accurately locate features (waste pits, truncated material layers along the mesa, fence boundaries, and the compliance boundary), and an incorporation of a recent three-dimensional geologic framework to define the hydrogeologic layers.

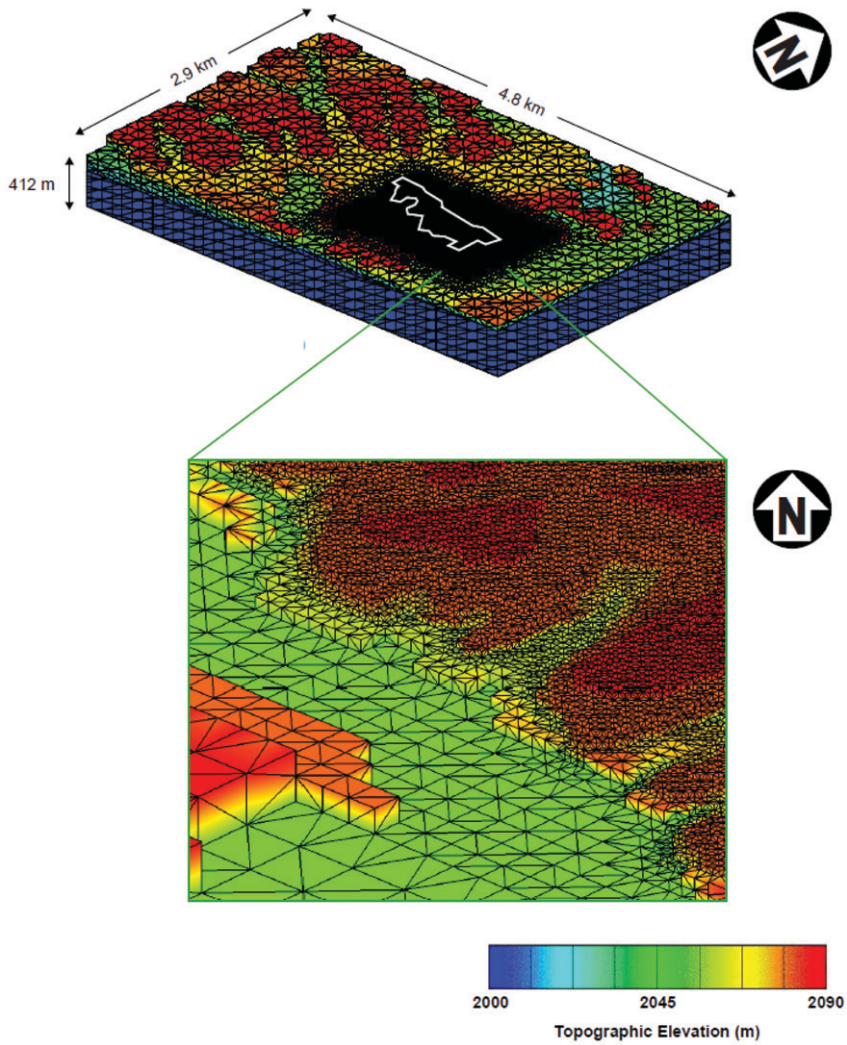


Fig. 3. Computational mesh used to calculate breakthrough curves (BTCs) from a release near the center of Material Disposal Area (MDA) G, outlined in white on the top figure. The compliance boundary at which we analyzed particle breakthrough is a north–south plane located 100 m to the east of the farthest east point of the MDA G boundary.

The computational mesh incorporates a refinement technique that provides high resolution near the MDA G waste pits and shafts and lower resolution away from the MDA G fence line (Fig. 3). The grid measures 4750 m from east to west and 2875 m from north to south, with an area of nearly 15 km^2, which is large enough to avoid boundary effects. The surface elevation of the grid was interpolated from a high-resolution DEM (Carey and Cole, 2002) that ranges from 2150 m asl in the northwest to approximately 2000 m (6560 ft) asl in the southeast. The grid extends well below the water table and is used to follow contaminant pathways from the surface of the disposal facility through the vadose zone, into the saturated zone, and finally to the compliance boundary. The grid spacing between nodes in the horizontal direction reaches a minimum of 7.8 m in the vicinity of the pits and shafts and is coarsest (125 m) in the regions farthest from MDA G. The high-resolution section extends >100 m beyond the boundary of MDA G to ensure that lateral transport issues can be adequately addressed. More details on our mesh generation techniques can be found in Miller et al. (2007).

Initial and Boundary Conditions

All lateral boundaries in the vadose zone were assumed to be no flow. Lateral gradients on these boundaries were not considered because previous modeling studies of the Pajarito Plateau found the magnitude of lateral gradients in the unsaturated zone to be generally quite small (Birdsell et al., 2000; Stauffer et al., 2005b).

Groundwater flow in all simulations was assumed to be from west to east following the water table gradient in the area. The gradient was fixed for all simulations and was based on a water table elevation of 1798 m along the western boundary and 1737 m to the east. These elevations yield an average gradient across the domain of approximately 0.013 m/m toward the Rio Grande. This gradient, based on data from Stone et al. (1999) and Keating et al. (2003), is expected

to capture the general trend of flow near the water table. The northern and southern boundaries in the saturated zone are no flow.

Because the goal of this study was to quantify the uncertainty due to the hydrologic properties of the stratigraphic units, we fixed the infiltration flux as a deterministic parameter to 1 L/(m^2 yr) = 1 mm/yr (Stauffer, 2006), a common sub-root-zone value for the very dry mesa tops of the Pajarito Plateau (Stauffer et al., 2005b)

Hydrologic Properties of Unsaturated Soils

To describe unsaturated flow, the constitutive relationships of saturation and capillary pressure must be specified. In this study, we adopted the van Genuchten model (van Genuchten, 1980):

$$P_{\text{cap}} = \frac{1}{\alpha}\left[\hat{S}^{-(1/m)} - 1\right]^{1/n} \quad [1]$$

$$\hat{S} = \frac{S - S_r}{S_{\max} - S_r} \quad [2]$$

where P_{cap} is capillary pressure, $\hat{S}$ is the effective saturation, S is water saturation, $S_{\max}$ and S_r are the maximum and residual saturation, respectively, (S_r can be converted from the saturated water content θ_s and is the residual [irreducible] water content θ_r), α is a pore-size distribution parameter, n is a fitting parameter, and $m = 1 - 1/n$. In the formulation used within FEHM, the van Genuchten–Mualem l parameter (related to pore connectivity) is fixed to 0.5 and we did not attempt to explore uncertainty with respect to this parameter in the current analysis.

Springer (2005) examined geographic differences among vadose-zone hydrologic properties across LANL as a means of estimating vadose-zone model parameters for the Bandelier Tuff. These hydrologic properties included bulk density, saturated water content, saturated hydraulic conductivity, fitted parameters such as the van Genuchten parameters (α and n), and residual water content. Nonparametric analyses were used to identify differences in the measured hydrologic properties for different lithologic units within various LANL technical areas and mesa-top vs. canyon settings. Hydrologic properties for Tshirege Units 1v and 1 g were essentially the same.

Hydrologic properties in four of these units were considered as random variables following some particular distributions inferred from the data. The saturated hydraulic conductivity K_s, the pore size distribution parameter α, fitting parameter n, saturated water content θ_s, and residual water content θ_r were treated as piecewise random functions, which varied from unit to unit but were constants in each unit. In this study, K_s and α were assumed to follow lognormal distributions, n was assumed to follow a truncated normal distribution (with a lower bound of 1.0), and θ_s and θ_r to follow normal distributions. In total there were five random variables sampled for each of four geologic units considered in this study. The particular distributional assumptions made in this study are consistent with the findings of Springer (2005) based on field data at the Los Alamos site (Table 1). It is noted from the table that rock properties at the site have moderate variability. In particular, moderate variability of the pore-size distribution parameter α and fitting parameter n will have a significant effect on flow and solute transport (Lu and Zhang, 2002).

Model Approach

Generation of Latin Hypercube Samples

The realizations of rock properties were generated using the LHS method. This technique was first described by McKay et al. (1979) and was further elaborated by Iman et al. (1981). This sampling scheme is a form of stratified sampling that can be applied to multiple variables. The method is commonly used to reduce the number of runs necessary for a Monte Carlo simulation to achieve a reasonably accurate random distribution. When sampling a set of N variables, the range of each variable is divided into M equally probable intervals, where M is a predetermined number of samples needed. The M sample points are then placed to satisfy the Latin hypercube requirements, i.e., each column or row of the hypercube having one and only one sample point. One of the advantages is that this sampling scheme does not require more samples for more dimensions (variables).

Table 1. Parameter statistics for four stratigraphic units (modified from Springer, 2005).

Parameter	Unit	n	Mean	SD
Hydraulic conductivity (ln K), m/s	Qbt2	17	−13.12	1.13
	Qbt1v	44	−13.63	1.06
	Qbt1g	17	−13.59	0.69
	Qbof	12	−12.89	0.46
Pore-size distribution parameter (ln α), 1/m	Qbt2	8	−0.76	0.69
	Qbt1v	34	−0.87	0.94
	Qbt1g	9	−0.90	0.32
	Qbof	12	−0.53	0.21
Shape parameter n	Qbt2	8	2.06	0.51
	Qbt1v	34	1.73	0.28
	Qbt1g	9	1.81	0.17
	Qbof	12	1.76	0.25
Volumetric saturated water content (θ_s)	Qbt2	10	0.41	0.03
	Qbt1v	35	0.5	0.04
	Qbt1g	14	0.46	0.05
	Qbof	12	0.43	0.01
Volumetric residual water content (θ_r)	Qbt2	8	0.01	0.013
	Qbt1v	34	0.003	0.009
	Qbt1g	9	0.01	0.015
	Qbof	12	0.019	0.015

The number of required Monte Carlo simulations depends on the variability of the model parameters. For large-scale, three-dimensional flow and transport simulations, especially in unsaturated porous media, the computational cost in terms of CPU time for each simulation can be very high, partially because of nonlinearity of problems, and we may not be able to conduct thousands or even more of large-scale simulations. Unlike traditional sampling methods that take samples for all parameters from their probability spaces and then run model simulations with these samples directly, the proposed RCS method takes subsamples of these parameter realizations based on the retention curves to reduce the number of required realizations while retaining reasonable accuracy.

We generated four sets of LHS realizations, with 25, 50, 100, and 1000 divisions (also the number of realizations), respectively. First, the results from these sets were used for investigating convergence of the Monte Carlo simulations, i.e., the number of required realizations. We show below that 1000 realizations was sufficient for convergence, and we then discuss our transport behaviors based on the results from 1000 realizations. We also explored the efficiency of the RCS scheme by comparing the results from the first three sets of realizations with those from subsets of the 1000 LHS realizations.

In the example we present, LHS was done in a correlated manner for each of the four rock units. For example, in the first realization, we chose the first LHS values for each rock unit, in the second realization, the second for each rock unit, and so on. Because the LHS sampling scheme was identical for each rock unit, the first sample pulled for each rock unit was from approximately the same region of the five-dimensional sample space, while the final pick for each rock unit was also from the same region of the five-dimensional sample space. Thus the sets of LHS simulations were designed to span the sample space (i.e., low suction to high suction in all rocks) but not to include combinations of sampling that involved sets of properties from different regions of the sample space for individual rock units (i.e., low suction in one rock unit and high in another). In other words, in a given simulation, if one rock unit was sampled from the low suction end of its sample space, all other rocks types were sampled from the same low suction region of their sample spaces. This simplification was done to reduce the total number of runs for convergence in the five-dimensional space we were exploring for each of the four rock units of the example problem.

In a typical Latin hypercube Monte Carlo simulation, all Latin hypercube samples have to be used in the simulation because values sampled from parameter spaces depend on the number of needed samples. As a result, direct subsampling from these Latin hypercube realizations is not appropriate. Our hypothesis was that, using the RCS scheme, if a simulation problem only allowed us to conduct a small number (say, N) of simulations, generating a larger number of samples (M) and then subsampling N samples from M retention curves would outperform N samples directly generated from the LHS method. The generation of 1000 LHS retention curves can be done virtually instantaneously, while running 1000 LHS simulations could easily take many thousands of CPU hours to complete.

Retention Curve Subsampling Scheme

Instead of subsampling the five parameter fields for each rock unit, we took a subset of samples based on the retention curves for each rock unit with the following procedure:

1. Retention curves (P_{cap}) were computed for all M realizations using Eq. [1] for the range of saturation from 0 to 1. As an example, retention curves computed for four different units based on 1000 LHS samples are illustrated in Fig. 4, where the red and blue curves represent the 5th and 95th percentiles, and symbols are measurements computed from measured gravimetric moisture content and matric potential. It can be seen in the figure that retention curves for Qbt2 and Qbt1v have the largest spread because of the high variability in their rock properties. It is interesting to note that the spread of retention curves for Qbt1 g is smaller than that for Qbof, while the variability of rock properties for Qbt1 g is larger than that for Qbof except for the variability of the van Genuchten fitting parameter n, which indicates that the variability of n may have a largest impact on the degree of spread of retention curves. This is consistent with a previous stochastic analysis (Lu and Zhang, 2002). In addition, nearly all saturation measurements fall between the 5th and 95th percentiles. More importantly, the distribution of retention curves in any subplot of Fig. 4 is non-uniform, i.e., in some areas the retention curves are very crowded while in other areas they appear very sparse. More curves are distributed in the low matric potential area than in the high matric potential area. This important feature motivated us to subsample realizations based on retention curves.

2. Because of the nonlinear relationship between the capillary pressure and rock properties, retention curves (P_{cap} vs. saturation S) computed from all realizations of rock properties may cross over (Fig. 4). Therefore, these curves were sorted (and labeled) according to the capillary pressure at a given saturation value, say, $S = 0.5$. The impact of such a selected saturation value to the simulation results was investigated.

3. The M curves were divided into N groups (or intervals), where N is the number of desired realizations. It was expected that N would be substantially smaller than M, while the statistics computed from these N realizations could adequately approximate those computed from the original M realizations.

4. One realization from each interval was chosen randomly (i.e., partially random).

5. Sampling for the RCS method was correlated in the same manner as described for the LHS sampling, where all four rock units were sampled from the same region of the sample space for a given realization. Thus, for one realization, each rock unit could reach steady state at a wetter or drier value from its range of behavior, but wet–dry behavior is not juxtaposed.

Design of Numerical Experiments

Before demonstrating the effectiveness of the proposed RCS approach, we investigated the number of Monte Carlo simulations

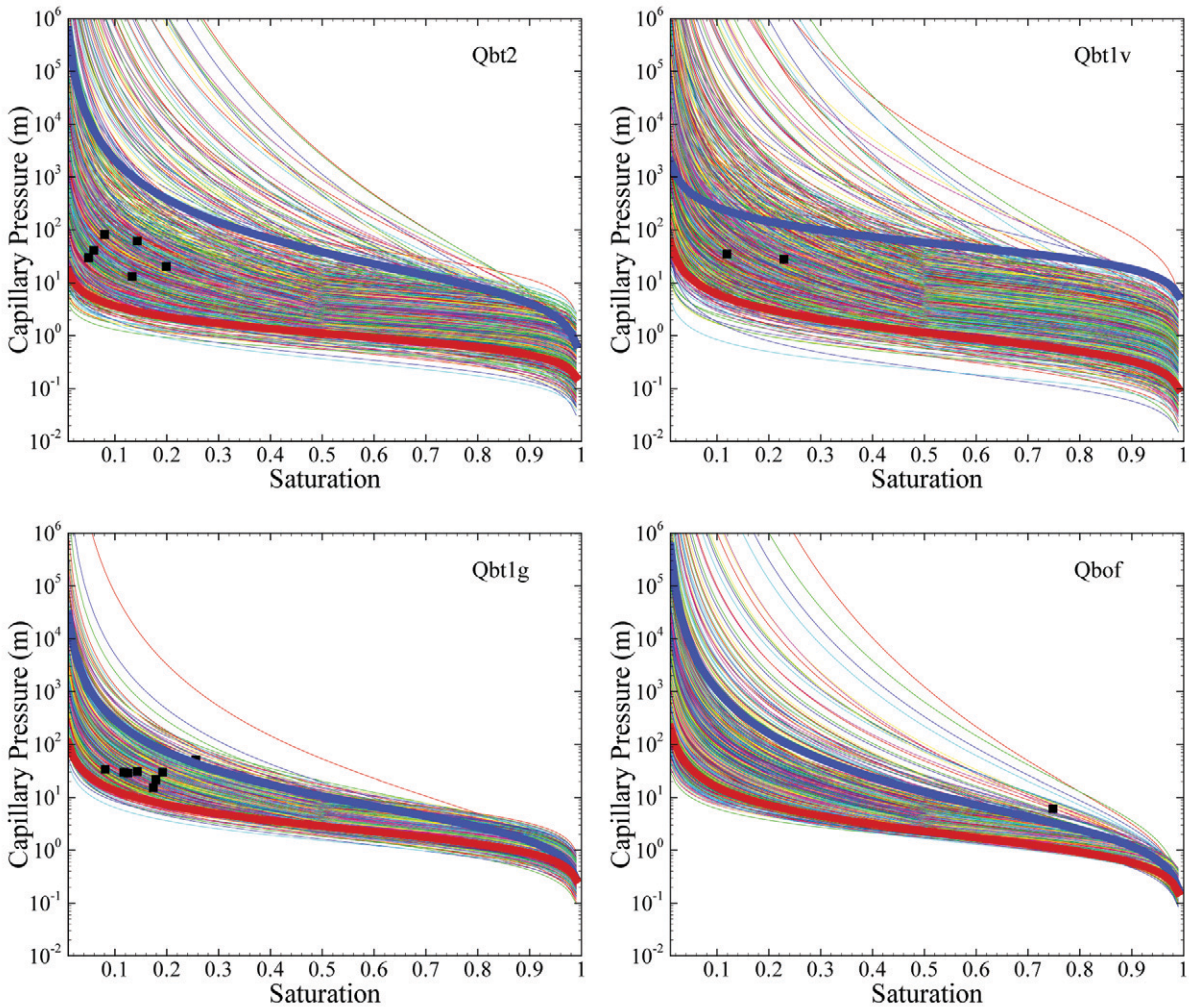


Fig. 4. Retention curves computed for four different stratigraphic units based on 1000 realizations of rock properties generated by the Latin hypercube sampling method. The red and blue curves represent the 5th and 95th percentiles sorted at saturation $S = 0.5$, and the black squares are measured data.

needed for the LHS realizations to converge to a stable answer. In each of the following cases, a number (M) of LHS realizations were generated and all realizations were used in Monte Carlo runs:

- Case LHS-25: $M = 25$
- Case LHS-50: $M = 50$
- Case LHS-100: $M = 100$
- Case LHS-1000: $M = 1000$

Once the convergence of the Monte Carlo simulations was understood, we compared the results with four cases with different RCS sizes, where N is the number of subsampled retention curves:

- Case RCS-25: $N = 25$
- Case RCS-50: $N = 50$
- Case RCS-100: $N = 100$
- Case RCS-200: $N = 200$

These subsamples were taken randomly from different intervals. For instance, for Case RCS-25, all 1000 LHS-generated retention curves for each of four stratigraphic units were divided into 25 intervals, sorted by suction at saturation $S = 0.5$, and one curve was randomly picked from each group of 40 retention curves that spanned the set of 1000 curves.

In addition to comparing with Case LHS-1000 for investigating the convergence of simulations in term of the LHS sample size, the results from Cases LHS 25 to LHS-100 were also compared with those from Cases RCS-25 to RCS-100, respectively, to study their performance at the same computational cost. For instance, both Cases LHS-25 and RCS-25 included 25 model runs with compatible computational cost, but the realizations in Case LHS-25 were generated directly from LHS while the realizations in Case RCS-25 were down-selected from 1000 realizations of LHS samples based on sorted retention curves using the RCS method. The difference between Cases LHS-25 and RCS-25, compared with the "true" solution from Case LHS-1000, reflects their performance, i.e., their closeness to the "true" solution at the similar computational cost.

We conducted two additional sets of model runs with slight variations of Cases RCS-25 to RCS-200 either with fixed samples from intervals of retention curves (referred to as fixed below) or with completely random samples from 1000 realizations (referred to as random samples below).

Particle Tracking

For each selected realization of retention curves, we first ran a steady-state flow simulation by solving Richards' equation with the realization of rock properties corresponding to the selected retention curves. A particle tracking method, described by Robinson et al. (2012) and implemented in FEHM, was used for transport simulation, and a total of 15,625 (= 25 × 25 × 25) particles were uniformly released in a 1-m cube near the center of MDA G just beneath a waste disposal pit with coordinates x = 501,304.4 m, y = 535,715.4 m, and z = 2024 m. The particle release was located in the unsaturated zone approximately 300 m above the regional water table. Dispersion and diffusion were ignored in this study. The number of particles reaching the compliance boundary was recorded, and normalized cumulative BTCs were derived from these records. All 1000 BTCs from Case LHS-1000 are illustrated in Fig. 5.

Results and Discussion

Convergence of Monte Carlo Simulations

The convergence of Monte Carlo simulations is, in general, assessed by visually examining the stability of the quantities of interest (say, mean concentration) as a function of the number of realizations. Ballio and Guadagnini (2004) proposed a methodology for convergence analysis of the first two moments (mean and variance) of a state variable of interest from Monte Carlo simulations. Certainly, the convergence criteria depend on the metrics used in the Monte Carlo analysis. For normalized cumulative BTCs, we chose four metrics: the 95th percentile of the curves (early-time BTC), the fifth percentiles (late-time BTC), the mean, and the median (50th percentile). Figure 6 compares these four sets of normalized BTCs derived from cases with different numbers of LHS realizations. Also compared in the figure is a BTC from transport modeling using the mean rock properties listed in Table 1. It should be noted that the BTCs derived from numerical transport simulations of different parameter realizations may have been recorded at different time discretizations, and therefore before evaluating the statistics of the BTCs, numerical interpolation may be needed so that all BTCs have values at the same time discretization. While the mean BTC was computed from the arithmetic average of BTCs, the 5th and 95th percentiles and median curves were determined in a different way because of crossover among these curves. For example, the 5th percentile curve in Fig. 6b for LHS-1000 was determined by first finding the 50th smallest value at each time out of 1000 values at that time and then connecting these values for all times. This 5th percentile curve may or may not coincide with any of the BTCs from the 1000 realizations.

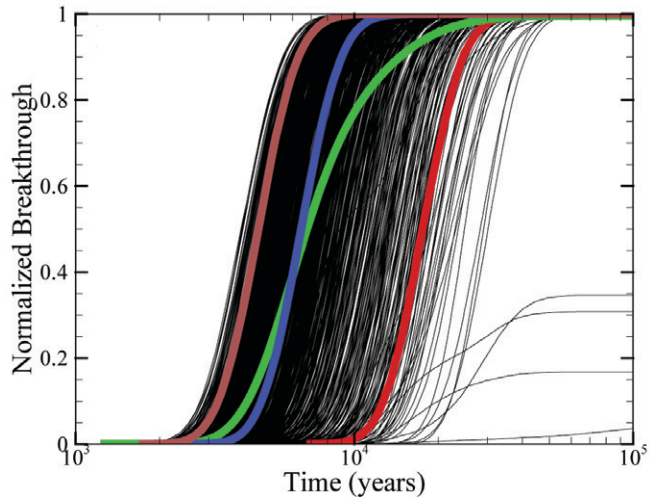


Fig. 5. Normalized cumulative breakthrough curves from 1000 Monte Carlo realizations and some particular curves: 5th percentile (red), mean (green), median (blue), and 95th percentile curve (brown).

Several observations can be made from this figure. First, Fig. 6 shows that the four quantities of interest (5th and 95th percentiles, mean, and median breakthroughs) from 100 realizations are very close to those from 1000 realizations, indicating that 100 LHS realizations are approaching convergence to the 1000 LHS Monte Carlo simulations. Excluding the 5th percentile (late-time BTC), the LHS appears to have been well converged by 50 simulations. It can be seen from the figure, however, that 25 LHS realizations did not converge to the "truth," with wide disparities in the 95th percentile, mean, and median values. Second, we note that the BTC derived from the mean rock properties is almost identical to the median BTC (Fig. 6d), while it is substantially different from the mean BTC (Fig. 6c), which was computed by arithmetic averaging of BTCs from all 1000 LHS realizations.

General Behaviors of Breakthrough Curves

We next discuss the general behavior of the BTCs based on 1000 LHS realizations of rock properties. As illustrated in Fig. 5, there is substantial variability in these BTCs. The difference among them is about one order of magnitude. Another interesting observation is that the curves are denser at earlier than later times. If the distribution of travel time is plotted for a given breakthrough level (say, half of the mass reaching the water table), it is expected that the distribution would have a long tail. A similar conclusion can also be made from the positions of the 5th, 50th, and 95th percentile curves. The distance between the 95th and 50th percentile curves is much smaller than the distance between the 50th and 5th percentile curves, which means that there is a higher curve density between the 95th and 50th percentiles than between the 50th and 5th percentiles.

One important feature of these percentile curves is that they all resemble BTCs from individual realizations, while the mean curve computed by arithmetic averaging of all 1000 realizations deviates

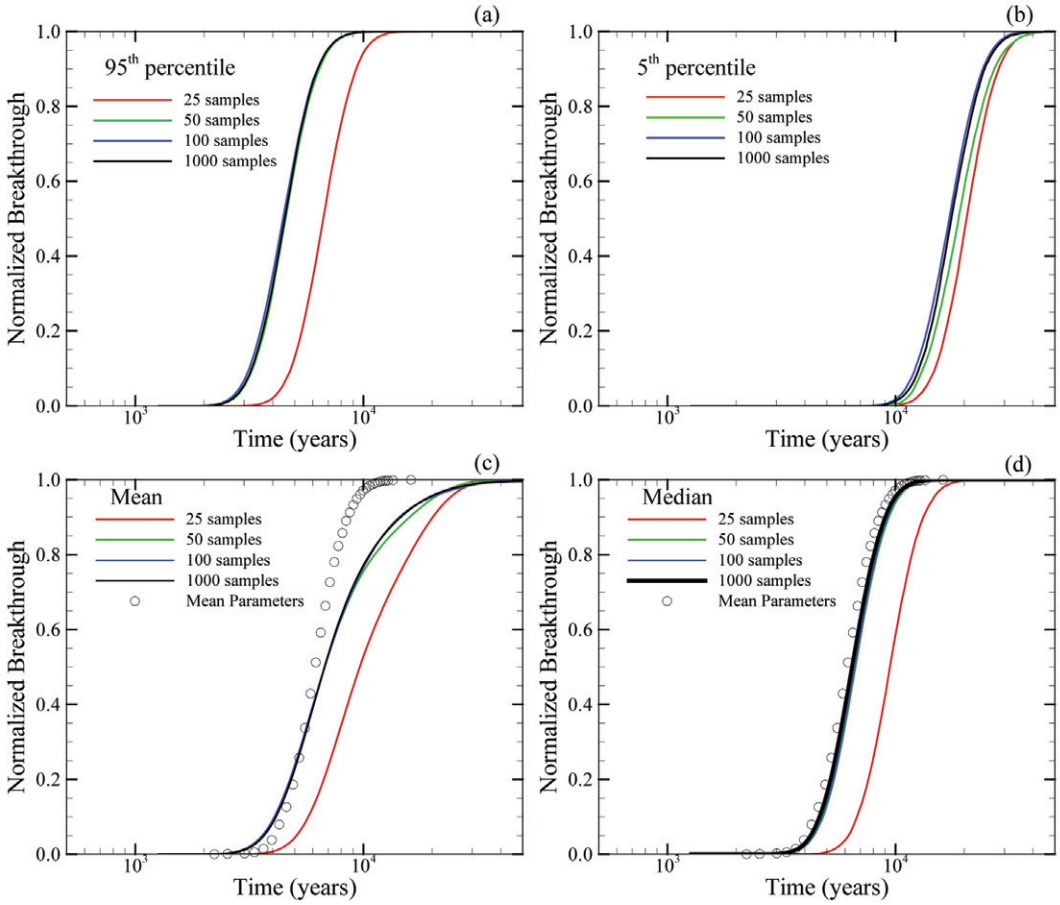


Fig. 6. Comparison of (a) 95th percentile, (b) 5th percentile, (c) mean, and (d) median of breakthrough curves derived from different numbers of Latin hypercube sampling realizations.

significantly from the individual BTCs. In general, the averaged curve may overestimate the mass flux at earlier and later times but underestimate it at intermediate time. More importantly, the averaged curve may underestimate the peak of the mass flux. This indicates that the mean curve from conventional averaging may not be a good estimate of the true mean behavior, and the 50th percentile curve may outperform the mean curve. From a statistical point of view, the conventional average of all sampled curves will be the best estimate if the L_2 norm is used to measure the distance between the mean and any individual sample. It is well known that the L_2 norm (square) is sensitive to outliers, and several approaches have been proposed to resolve this problem. The methods based on the L_1 norm (absolute value), which yields the median of the samples, are more robust to outliers than the methods based on the L_2 norm. An approach for finding a better mean curve was discussed further by Lu and Stauffer (2012).

Reduction of Monte Carlo Runs by Subsampling

We are now at a position to investigate how the RCS method could save computational effort. Figure 7 shows the quintiles of interest for the RCS method, and it is clear that with only 25 subsamples, the RCS method was able to capture the mean, median, and 95th percentile much more accurately than the LHS-25 case from Fig. 6. In all cases, both LHS and RCS, the 5th percentile did not converge until 100 samples, which probably is a result of the limited number of curves near the late-time tail of the distribution. Finding a method to more rapidly converge on the 5th percentile will require more analysis, but we believe a weighting scheme could be devised to address this problem from within the simpler RCS method.

Figure 8 compares the median of BTCs derived from both the LHS and RCS methods. The subsamples from 1000 realizations were taken in three different ways (taking Case RCS-25 as an example): fixed samples, i.e., the 20th curve from each group of 40 retention curves, random samples from intervals of 40 retention curves, and completely random samples taken from 1000 LHS realizations without taking into account the order of the retention curves.

It seems from Fig. 8 that there is almost no difference between subsampling randomly from predetermined intervals or fixed subsampling from these intervals, where the intervals were obtained by dividing all 1000 samples into N groups (N = the number of subsamples) based on sorted retention curves. All these cases with

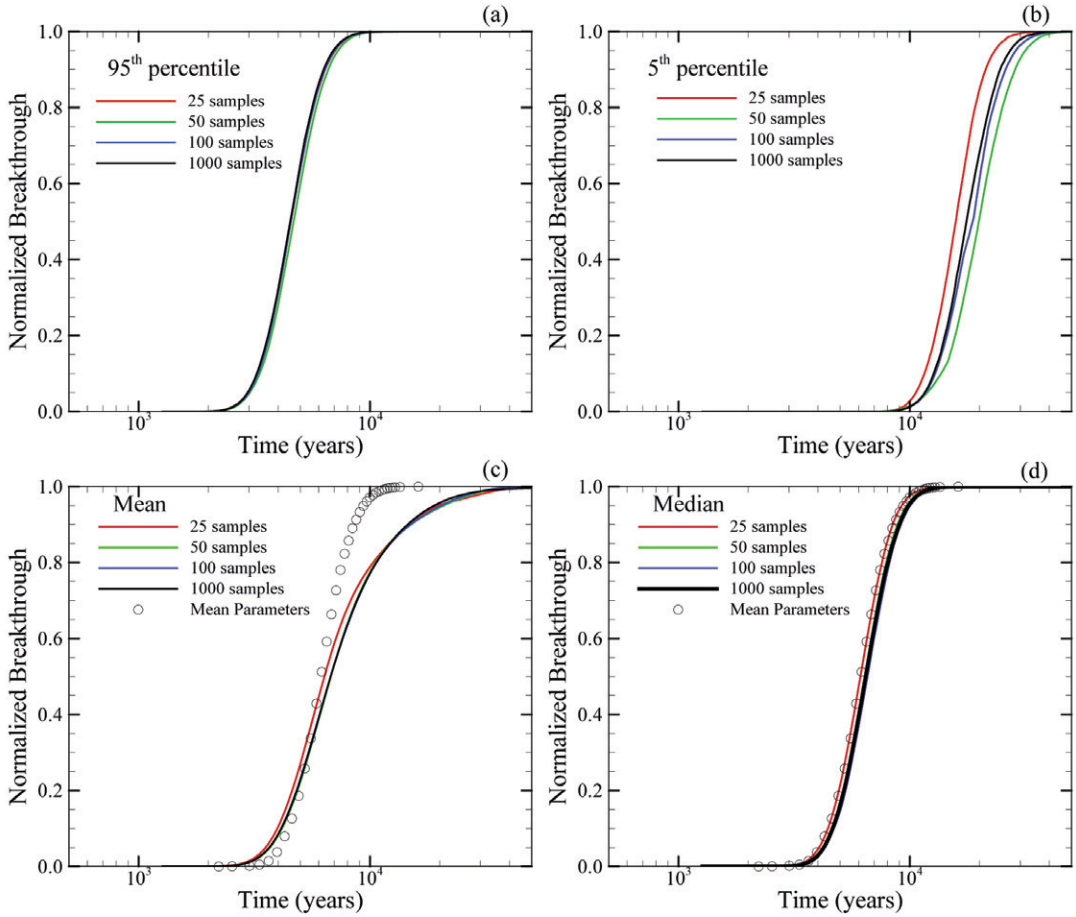


Fig. 7. Comparison of (a) 95th percentile, (b) 5th percentile, (c) mean, and (d) median of breakthrough curves derived from different numbers of retention curve samples using the retention curve subsampling method.

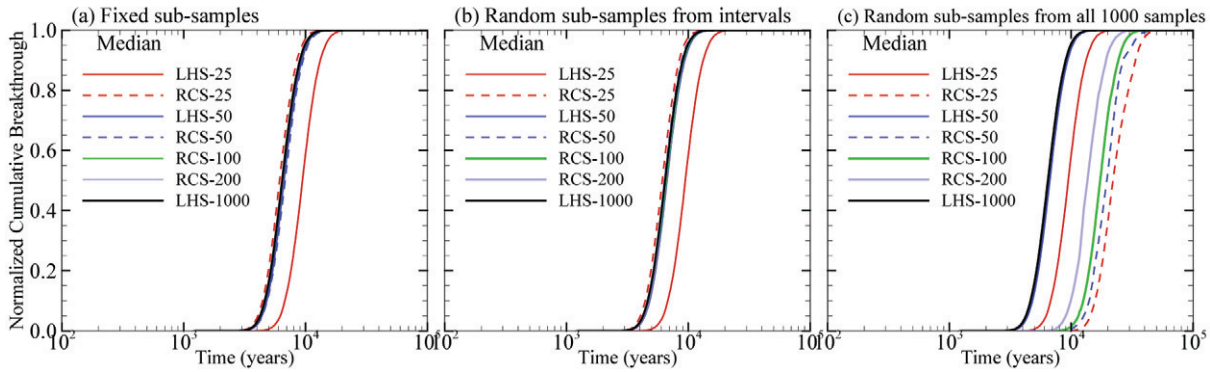


Fig. 8. Comparison of median of breakthrough curves derived from various Latin hypercube sampling (LHS) runs with those of retention curve subsampling (RCS) method subsamples taken from 1000 LHS retention curves in three different ways: (a) fixed samples from intervals of retention curves, (b) random samples from intervals of retention curves, and (c) completely random taken from 1000 LHS realizations without taking into account their retention curves.

subsampling from 1000 realizations yielded consistent results (Fig. 8a and 8b). Even the case with only 25 RCS subsamples provided excellent results. On the other hand, the results from the case with 25 LHS realizations deviated significantly from the "true" solutions. This clearly indicates that the proposed subsampling scheme has advantages over the commonly used approach that runs models directly using all LHS realizations of rock parameters. One probable explanation is that, while 25 LHS samples do cover the probability space of the rock properties, the retention curves computed from these 25 LHS samples do not provide an adequate representation of the probability space of the retention curves available through 1000 LHS-generated retention curves. It should be noted that such a computational advantage becomes even more significant if the variability of the rock properties increases.

If subsamples are taken randomly from 1000 samples without considering the order of the retention curves, however, as illustrated in Fig. 8c, the non-RCS scheme does not work well. The results from 25 LHS realizations are much closer to the "true" median curve than the results from as many as 200 random subsamples, which is expected because 200 subsamples taken randomly from 1000 realizations will probably not adequately cover the probability space of the retention curves.

Another way to assess the efficiency of sampling schemes is to evaluate how fast the estimated variance of BTCs converges to the "true" variance, which was estimated from 1000 LHS realizations. Figure 9 depicts the variance of BTCs derived from the range of LHS realizations with those of different numbers of RCS subsamples taken from 1000 LHS retention curves in the three different subsampling approaches. The figure further confirms that the subsampling method based on sorted retention curves in different intervals (Fig. 9a and 9b) is much better than resampling randomly from all 1000 realizations (Fig. 9c).

Another question is how the simulation results are affected by the saturation level at which the retention curves are sorted. Because of the crossover of retention curves, if we sort these curves at a different saturation level, the order of the curves will be different, and therefore the selected curves from the subsampling procedure will also be different. The question is: How will differences in selected curves affect the simulation results? To answer this question, we performed a similar set of simulations, but the retention curves were sorted at a saturation value of 0.25 instead of 0.5. The median and the variance of BTCs from both saturation values are compared in Fig. 10. This figure indicates that the impact of sorting retention curves using different saturation values is very small. We suggest, however, that sorting be performed where the curves are better behaved because, toward the wet and dry ends of Fig. 4, the curves tend to be less well behaved with more crossover, and sorting in these regions may lead to unexpected results.

Finally, the decision to sort the retention curves was based on the idea that many non-unique combinations of retention parameters can lead to quite similar curves, something that is not taken into consideration in a standard LHS search of the parameter space. This idea was based on the work of Stauffer et al. (2009), whose results showed that many combinations of transport properties could lead to realizations that yielded nearly identical results. Thus, our methodology reduces the total number of realizations needed to span the range of system behavior by first plotting the 1000 LHS samples, then selecting from this group to reduce duplication of the curves that lie along the same suction vs. saturation trajectory. Our analysis was not exhaustive and we plan to explore in more detail how such a sampling scheme can be improved. For example, plotting the 25 subsampled curves from the suction vs. saturation space with suction as a function of the relative permeability space shows that they do span the range of behavior but not with the same regular spacing that was imposed on the original set of subsamples.

This discussion demonstrates the validity of our hypothesis: To reduce the computational burden of LHS techniques when dealing with uncertainty in retention curves, a large number LHS retention curves may be generated, the curves then sorted by suction at a given saturation (0.75–0.25), and a small subset chosen using the RCS method. The robust nature of the results suggests that the RCS method maybe be applied to a range of simulations where uncertainty in retention plays a strong role in BTC behavior.

Impact of Saturation on Transport

Above, we investigated BTCs observed at the proposed MDA G compliance boundary. We next explore the details of solute transport in three-dimensional space. In particular, we are interested in the impact of saturation on particle transport behaviors.

Figure 11 compares the first 10 particle trajectories (out of 15,625) for three realizations corresponding to three different retention curves: realizations 100, 500, and 900 (labeled as R100, R500, and

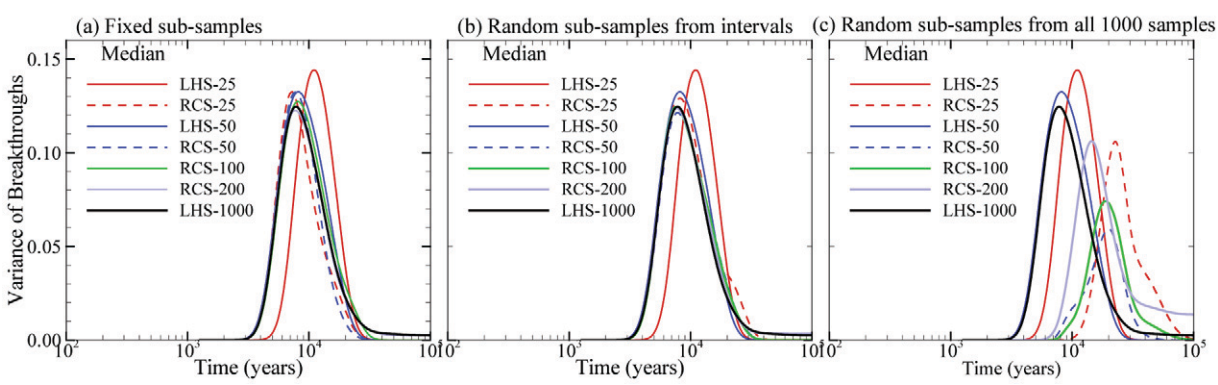


Fig. 9. Comparison of variance of breakthrough curves derived from Latin hypercube sampling (LHS) realizations with those of retention curve subsampling (RCS) method subsamples taken from 1000 LHS retention curves in three different ways: (a) fixed samples from intervals of retention curves, (b) random samples from intervals of retention curves, and (c) completely random taken from 1000 LHS realizations without taking into account their retention curves.

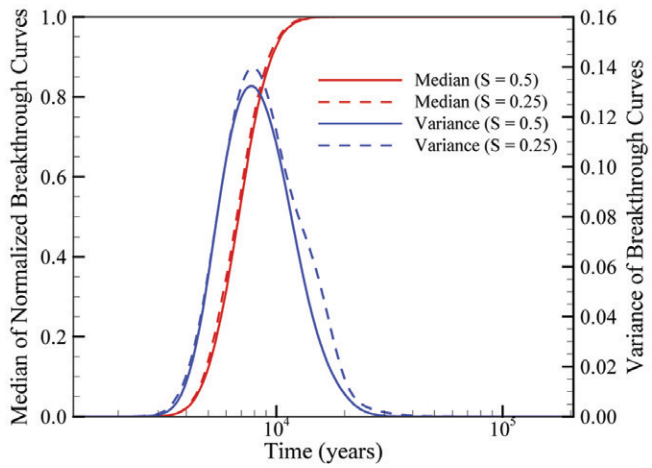


Fig. 10. Comparison of median and variance derived from retention curves sorted at saturation $S = 0.25$ and 0.5.

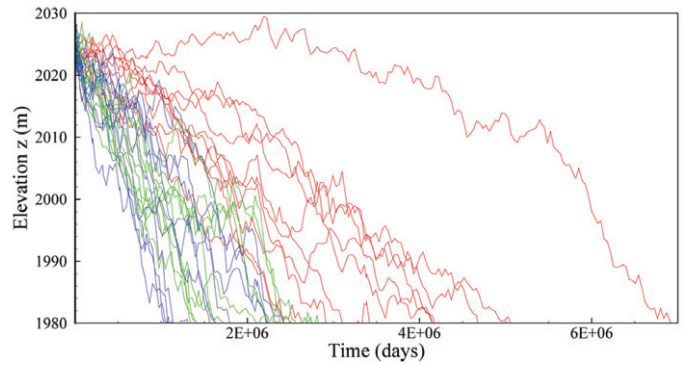


Fig. 11. Selected particle trajectories for three realizations: R100 (red), R500 (green), and R900 (blue), representing the degree of saturation from low to high.

R900), where the realizations are numbered based on the sorted retention curves at the saturation value of 0.5. Realization R100 has the largest capillary pressure (i.e., suction) at $S = 0.5$ among the three, R900 has the smallest suction, and R500 intermediate suction. It is evident from the figure that the variability of these trajectories is large when the capillary pressure is large (red curves). In addition, particles move faster when the capillary pressure is small (blue curves). Figure 12 compares vertical profiles of the capillary pressure and saturation along the particles' release location for the above three realizations. The zigzags of the curves in the figure are due to the layering of the stratigraphic units. This figure shows that, although the capillary pressure was highest for R100, the saturation for this realization was also highest among the three. This can also be seen from Fig. 13, a scatter plot showing the positive correlation between capillary pressure and saturation, with the highest saturation for R100 and lowest saturation for R900. This somewhat counterintuitive result occurred because we had a fixed infiltration flux (q) at the top boundary [1 L/(m^3 yr)], and because of varying retention characteristics, the higher suction materials had to reach higher saturations before their relative permeability was such that they could transmit the required flux. This impacted particle transport through the true fluid velocity (defined as $u = q/\theta$, where θ is the effective water content) because at a fixed flux, particles move more quickly through rocks with lower water content.

The moments of particle trajectories computed from all 15,625 trajectories for these three moisture conditions are shown in Fig. 14. For all three cases, the mean position in the y coordinate is constant with time (Fig. 14a), indicating that the particles move in the x–z plane on average. Figure 14a also shows that particles moves nearly vertically at earlier time (up to 1.5×10^6 yr) and then start to move both vertically and laterally (in the x direction). The amount of lateral shift depends on the moisture condition; more lateral shift is observed with the increase of saturation from R900 to R100. Furthermore, the figure again indicates that particles move much faster under drier conditions (R900) than wetter conditions (R100). The spreading of particles in x and y coordinates under different moisture conditions is illustrated in Fig. 14b. It can be seen that lateral spreading is more noticeable when the medium is wetter (R100) due to slow vertical movement under wetter conditions, which allows particles to spread laterally. This

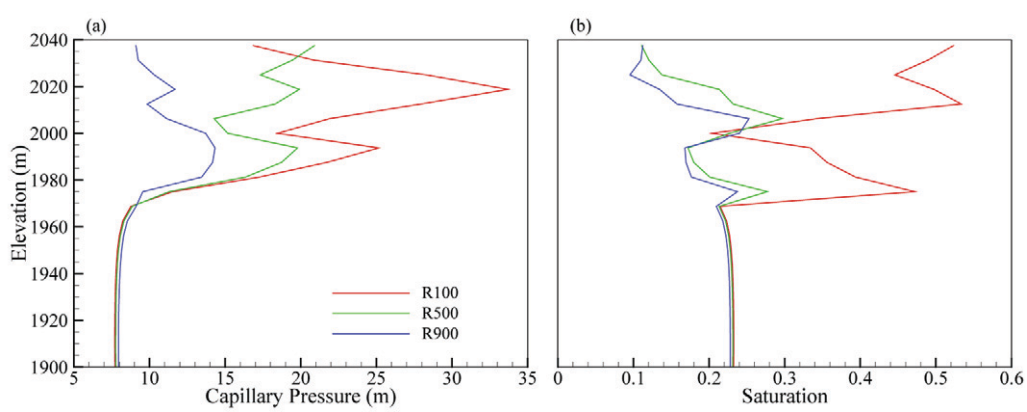


Fig. 12. Comparison of (a) capillary pressure and (b) saturation for three realizations—R100, R500, and R900—on the vertical profile passing through the particles' release location.

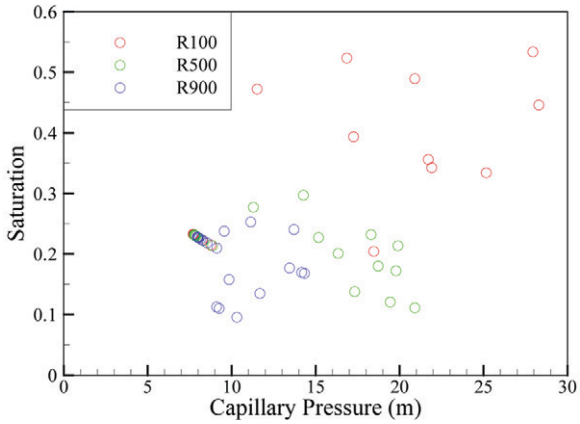


Fig. 13. Scatter plot showing the correlation between capillary pressure and saturation for three realizations.

can also be seen from Fig. 15, which shows the spatial distribution of particles at an elevation of z = 2000 m derived from superposition of all time steps.

Conclusions

In this study, we described a new approach for conducting Monte Carlo simulations of flow and transport in unsaturated heterogeneous porous media with piecewise random hydrologic parameters. The new method uses RCS of a large number of LHS-generated retention curves to reduce the number of simulations necessary to converge on the mean, median, and 95th percentile (early BTCs). Numerical analysis indicated that it is possible to reduce the number of model runs by a factor of two in the example problem presented by generating a large number of realizations of rock properties, calculating retention curves using selected constitutive relations (the van Genuchten model in this study), selecting a small number of retention curves from the retention curves sorted by suction at a fixed saturation, and running simulations using the realizations of rock properties corresponding to these selected retention curves. The saturation value at which the retention curves were sorted did not have a significant effect on the model results, but visual examination of 1000 retention

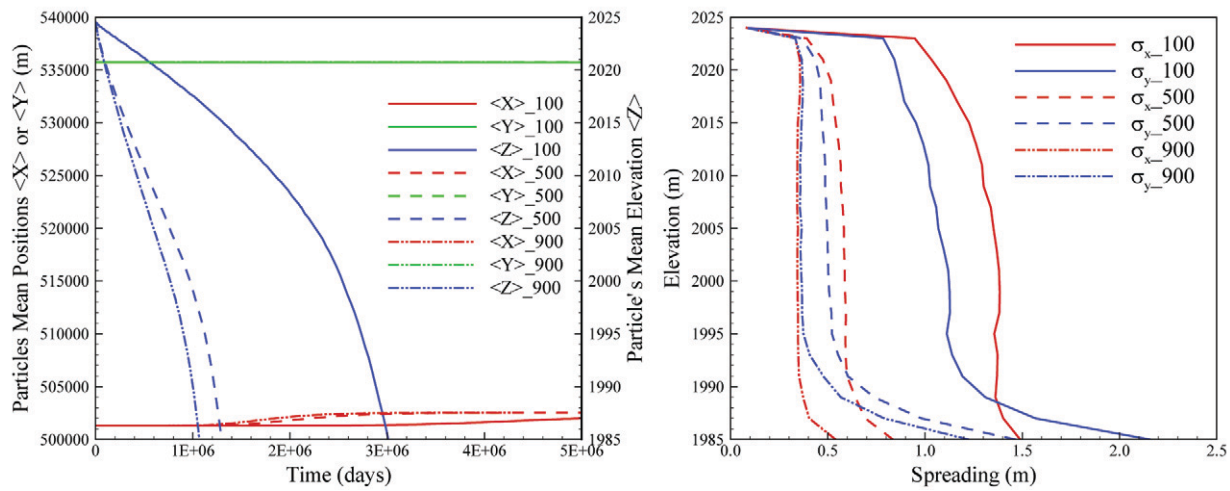


Fig. 14. Comparison of particles' (a) mean positions and (b) spreading for three moisture conditions ranging from dry (900) to wet (100).

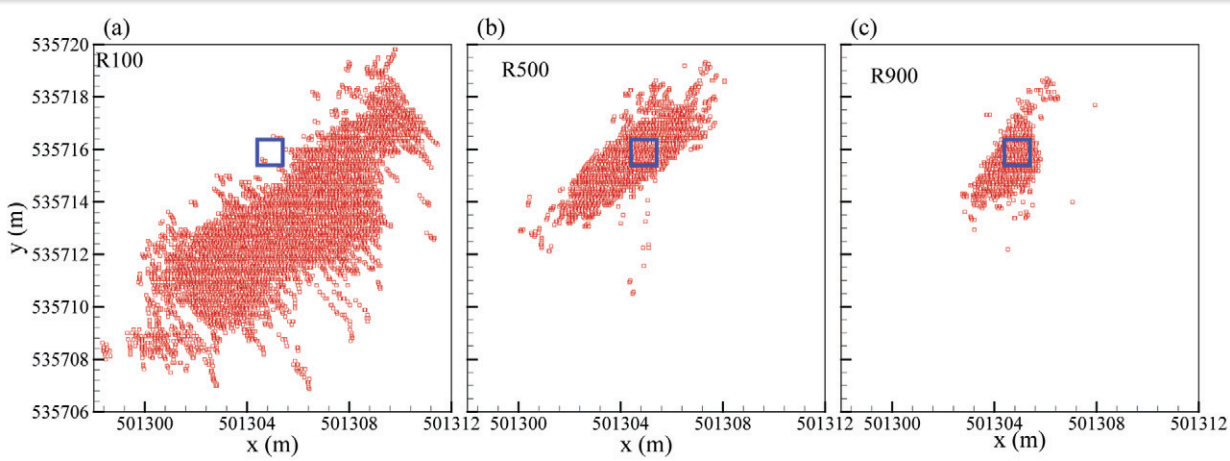


Fig. 15. Spatial distribution of particles when they cross a horizontal plane with elevation of 2000 m for three realizations: (a) R100, (b) R500, and (c) R900. Particles were released 24 m above this plane at coordinate x = 501,304.4 m, y = 535,715.4 m, z = 2024 m, marked with a blue square.

curves suggested that sorting be confined to between 25 and 75% saturation. In all cases, for both LHS and RCS, the 5th percentile did not converge until 100 samples, which probably was a result of the limited number of curves near the late-time tail of the distribution, and finding solutions to this issue is an ongoing area of research.

Finally, although the arithmetic mean of BTCs has been widely used to characterize mean transport behaviors, our example shows that the arithmetic mean of curves does not assemble to any BTC realization and overestimates the mass flux at earlier time but underestimates mass flux at later time. Our study shows that the median of the BTCs is more meaningful than the arithmetic mean of the curves. For the case studied, the median of the BTCs happened to be nearly identical to the BTC derived from the mean rock properties.

Acknowledgments

This study was supported by the Performance Assessment for MDA G, Los Alamos National Laboratory, under funding from the U.S. Department of Energy. We thank Rob Shuman for helpful review of the material. We also gratefully acknowledge the thoughtful reviews of two anonymous reviewers.

References

Ballio, F., and A. Guadagnini. 2004. Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. Water Resour. Res. 40:W04603. doi:10.1029/2003WR002876

Birdsell, K.H., A.V. Wolfsberg, D. Hollis, T.A. Cherry, and K.M. Bower. 2000. Groundwater flow and radionuclide transport calculations for a performance assessment of a low-level waste site. J. Contam. Hydrol. 46:99–129. doi:10.1016/S0169-7722(00)00129-7

Broxton, D.E., and D.T. Vaniman. 2005. Geologic framework of a groundwater system on the margin of a rift basin, Pajarito Plateau, north-central New Mexico. Vadose Zone J. 4:522–550. doi:10.2136/vzj2004.0073

Carey, J.W., and G. Cole. 2002. Description of the Cerro Grande Fire laser-altimetry (LIDAR) data set. Rep. LA-13892-MS. Los Alamos Natl. Lab., Los Alamos, NM.

Dash, Z.V. 2003. Validation test plan for the FEHM application version 2.21. Doc. 10086-VTP-2.20-00. Los Alamos Natl. Lab., Los Alamos, NM.

Hollis, D., E. Vold, R. Shuman, K. Birdsell, K. Bower, W. Hansen, et al. 1997. Performance assessment and composite analysis for the Los Alamos National Laboratory Disposal Area G. Rep. LA-UR-97-85. Los Alamos Natl. Lab., Los Alamos, NM.

Iman, R.L., J.C. Helton, and J.E. Campbell. 1981. An approach to sensitivity analysis of computer models: 1. Introduction, input variable selection and preliminary variable assessment. J. Qual. Technol. 13(3):174–183.

Keating, E.H., V.V. Vesselinov, E. Kwicklis, and Z. Lu. 2003. Coupling basin- and site-scale inverse models of the Espanola Aquifer. Ground Water 41:200–211. doi:10.1111/j.1745-6584.2003.tb02583.x

Krier, D., R. Longmire, and H.J. Turin. 1996. Geologic, geohydrologic and geochemical data summary of MDA G, TA-54, Los Alamos National Laboratory. Rep. LA-UR-95-2696. Los Alamos Natl. Lab., Los Alamos, NM.

Levitt, D.G. 2011. Modeling the movement of transient moisture through disposal units at Los Alamos National Laboratory Technical Area 54, Area G. Rep. LA-UR-11-05424. Los Alamos Natl. Lab., Los Alamos, NM.

Lu, Z., and P.H. Stauffer. 2012. On estimating functional average breakthrough curve using time-warping technique and perturbation approach. Water Resour. Res. 58:(in press). doi:10.1029/2011WR011506

Lu, Z., and D. Zhang. 2002. Stochastic analysis of transient flow in heterogeneous variably saturated porous media: The van Genuchten–Mualem constitutive model. Vadose Zone J. 1:137–149.

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

Miller, T.A., V.V. Vessilinov, P.H. Stauffer, K.H. Birdsell, and C.W. Gable. 2007. Integration of geologic frameworks in meshing and setup of computational hydrogeologic models, Pajarito Plateau, New Mexico. In: B.S. Kues et al., editors, Geology of the Jemez Mountains Region II. Fall Field Conference Guide Book 58. New Mexico Geol. Soc., Socorro. p. 492–500.

Painter, S., V. Cvetkovic, and D.R. Turner. 2001. Effect of heterogeneity on radionuclide retardation in the alluvial aquifer near Yucca Mountain, Nevada. Ground Water 39:326–338.

Riva, M., A. Guadagnini, D.F. Garcia, X. Sanchez-Vila, and T. Ptak. 2008. Relative importance of geostatistical and transport models in describing heavily tailed breakthrough curves at the Lauswiesen site. J. Contam. Hydrol. 101:1–13. doi:10.1016/j.jconhyd.2008.07.004

Robinson, A.B., S. Chu, and Z. Lu. 2012. Simulation of radionuclide transport through unsaturated, fractured rock: Application to Yucca Mountain, Nevada. Vadose Zone J. 11. doi:10.2136/vzj2011.0133

Rogers, D.B., and B.M. Gallaher. 1995. The unsaturated hydraulic characteristics of the Bandelier Tuff. Rep. LA-12968-MS. Los Alamos Natl. Lab., Los Alamos, NM.

Springer, E.P. 2005. Statistical exploration of matrix hydrologic properties for the Bandelier Tuff, Los Alamos, New Mexico. Vadose Zone J. 4:505–521. doi:10.2136/vzj2004.0067

Stauffer, P.H. 2006. Flux flummoxed: A proposal for consistent usage. Ground Water 44:125–128. doi:10.1111/j.1745-6584.2006.00197.x

Stauffer, P.H., K.H. Birdsell, M.S. Witkowski, and J.K. Hopkins. 2005a. Vadose zone transport of 1,1,1-trichloroethane: Conceptual model validation through numerical simulation. Vadose Zone J. 4:760–773. doi:10.2136/vzj2004.0120

Stauffer, P.H., and N.D. Rosenberg. 2000. Vapor phase transport at a hillside landfill. Environ. Eng. Geosci. 6:71–84.

Stauffer, P.H., and W.J. Stone. 2005. Surface water/groundwater connection at the Los Alamos Canyon weir site: 2. Modeling of tracer test results. Vadose Zone J. 4:718–728. doi:10.2136/vzj2004.0121

Stauffer, P.H., H.S. Viswanathan, B.A. Robinson, C.W. Gable, G.L. Cole, D.E. Broxton, et al. 2005b. Groundwater pathway model for the Los Alamos National Laboratory Technical Area 54, Material Disposal Area G. Rep. LA-UR-05-7393. Los Alamos Natl. Lab., Los Alamos, NM.

Stauffer, P.H., J.A. Vrugt, H.J. Turin, C.W. Gable, and W.E. Soll. 2009. Untangling diffusion from advection in unsaturated porous media: Experimental data, modeling, and parameter uncertainty assessment. Vadose Zone J. 8:510–522. doi:10.2136/vzj2008.0055

Stone, W.J., G.L. Cole, and M.A. Jones. 1999. Documentation for the preliminary hydrogeologic atlas for Los Alamos National Laboratory. Re. LA-UR-99-4619. Los Alamos Natl. Lab., Los Alamos, NM.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44:892–898. doi:10.2136/sssaj1980.03615995004400050002x

Vaniman, D., G. Cole, J. Gardner, J. Conaway, D. Broxton, S. Reneau, et al. 1996. Development of a site-wide geologic model for Los Alamos National Laboratory. Rep. LA-UR-96-2059. Los Alamos Natl. Lab., Los Alamos, NM.

Viswanathan, H.S., B.A. Robinson, A.J. Valocchi, and I.R. Triay. 1998. A reactive transport model of neptunium migration from the potential repository at Yucca Mountain. J. Hydrol. 209:251–280. doi:10.1016/S0022-1694(98)00122-X

Vrugt, J.A., P.H. Stauffer, T. Wöhling, B.A. Robinson, and V.V. Vesselinov. 2008. Inverse modeling of subsurface flow and transport properties: A review with new developments. Vadose Zone J. 7:843–864. doi:10.2136/vzj2007.0078

Zhang, Y., D.A. Benson, and B. Baeumer. 2007. Predicting the tails of breakthrough curves in regional-scale alluvial systems. Ground Water 45:473–484. doi:10.1111/j.1745-6584.2007.00320.x

Zyvoloski, G.A. 2007. FEHM: A control volume finite element code for simulating subsurface multi-phase multi-fluid heat and mass transfer. Rep. LA-UR-07-3359. Los Alamos Natl. Lab., Los Alamos, NM.

Zyvoloski, G.A., B.A. Robinson, Z.V. Dash, and L.L. Trease. 1997. Summary of the models and methods for the FEHM application: A finite element heat- and mass-transfer code. LA-13307-MS. Los Alamos Natl. Lab., Los Alamos, NM.