Percolation-Theory and Fuzzy Rule-Based Probability Estimation of Fault Leakage at Geologic Carbon Sequestration Sites

Yingqi Zhang, Curtis M. Oldenburg, and Stefan Finsterle

Earth Sciences Division, Lawrence Berkeley National Laboratory, University of California, MS 90R1116, 1 Cyclotron Road, Berkeley, CA 94720-8126, USA
e-mail: yqzhang@lbl.gov

Abstract. Leakage of CO_2 and displaced brine from geologic carbon sequestration (GCS) sites into potable groundwater or to the near-surface environment is a primary concern for safety and effectiveness of GCS. The focus of this study is on the estimation of the probability of CO_2 leakage along conduits such as faults and fractures. This probability is controlled by (1) the probability that the CO_2 plume encounters a conductive fault that could serve as a conduit for CO_2 to leak through the sealing formation, and (2) the probability that the conductive fault(s) intersected by the CO_2 plume are connected to other conductive faults in such a way that a connected flow path is formed to allow CO_2 to leak to environmental resources that may be impacted by leakage. This work is designed to fit into the certification framework for geological CO_2 storage, which represents vulnerable resources such as potable groundwater, health and safety, and the near-surface environment as discrete ‘‘compartments.’’ The method we propose for calculating the probability of the network of conduits intersecting the CO_2 plume and one or more compartments includes four steps: (1) assuming that a random network of conduits follows a power-law distribution, a critical conduit density is calculated based on percolation theory; for densities sufficiently smaller than this critical density, the leakage probability is zero; (2) for systems with a conduit density around or above the critical density, we perform a Monte Carlo simulation, generating realizations of conduit networks to determine the leakage probability of the CO_2 plume (Pleak) for different conduit length distributions, densities and CO_2 plume sizes; (3) from the results of Step 2, we construct fuzzy rules to relate Pleak to system characteristics such as system size, CO_2 plume size, and parameters describing conduit length distribution and uncertainty; (4) finally, we determine the CO_2 leakage probability for a given system using fuzzy rules. The method can be extended to apply to brine leakage risk by using the size of the pressure perturbation above some cutoff value as the effective plume size. The proposed method provides a quick way of estimating the probability of CO_2 or brine leaking into a compartment for evaluation of GCS leakage risk. In addition, the proposed method incorporates the uncertainty in the system parameters and provides the uncertainty range of the estimated probability.

Keywords: risk assessment, faults and fractures, fuzzy logic

Introduction

Large-scale injection of CO_2 into geologic formations is considered a potential mitigation method to reduce greenhouse gas emissions. The safety and effectiveness of geologic carbon sequestration (GCS) are achieved when injected CO_2 remains contained within the storage reservoir. Trapping mechanisms reducing the mobile CO_2 that could impact health, safety, or the environment include structural trapping, residual phase trapping, solubility trapping, and mineral trapping (IPCC 2005). Despite these, it is possible in some cases that CO_2 could unexpectedly leak upwards due to (1) the large amount of CO_2 injected and (2) the buoyant nature of CO_2 (Oldenburg et al. 2008). On the other hand, it is important to recognize that CO_2 is non-hazardous unless

concentrations are above certain levels. The key to public acceptance and success of GCS is to address leakage concerns, and to demonstrate that leakage risks are acceptably small.

Similar to what has been used for the risk assessment of nuclear waste repositories, general probabilistic theory and a features, events, and processes (FEP) scenario approach (Savage et al. 2004; Wildenborg et al. 2004, 2005) have been used to evaluate risks related to GCS. The FEP approach includes identifying all relevant FEPs, defining scenarios, and modeling environmental impacts and consequences. In this approach, the probability that certain FEPs will occur is usually assigned as input. Bowden and Rigg (2004) used a performance index to quantitatively characterize risks, where the likelihood and duration of each risk is determined by an expert panel and entered as input to each risk event. To date there has not been a quantitative risk evaluation based on geology and CO_2 plume characteristics for a given CO_2 storage site.

Three potential pathways may be available for CO_2 to escape from the storage formation to regions with vulnerable resources (e.g., a drinking water aquifer) (Espie 2004; Pruess 2008; Zweigel et al. 2004): (1) leakage through the caprock, (2) leakage through subvertical faults or fracture zones, and (3) leakage through abandoned wells. The focus of this study is on the second pathway—leakage through faults or fracture zones. The assessment of risks includes evaluation of both the probability that the CO_2 plume will reach drinking water or another vulnerable resource through a conduit, and the impact of the leakage, which requires knowledge of CO_2 flux, concentration, and total amount. The focus of our study is on the estimation of leakage probability (P_{leak}).

This work is designed to fit into the certification framework (CF) for geological CO_2 storage. The overall objective of the CF is to develop a simple framework for evaluating leakage risk for certifying operation and abandonment of geological CO_2 storage sites. In the CF, the term "compartment" is defined as a region containing vulnerable resources that could potentially be impacted such as potable groundwater and the near-surface environment. The locations of compartments are abstract in the sense that they may include disconnected pieces (Oldenburg et al. 2008). The same concept is used in this work as well.

The fundamental uncertainties addressed by the approach described in this paper are schematically presented in Fig. 1. Starting from the injection horizon, CO_2 enters the storage region and migrates as controlled by pressure, buoyancy, permeability, and capillary effects. Because plume migration distances and the presence of faults and fractures may only be estimated roughly from site characterization data prior to actual injection, the first unknown is whether the CO_2 plume will intersect a conductive fault that could serve as a conduit for CO_2 to leak through the sealing formation. Probabilistic approaches are needed to quantify the likelihood of this event based on the size of the CO_2 plume, which is highly uncertain given the uncertain properties of the deep storage reservoir. The second unknown is whether the conductive fault(s) intersected by the CO_2 plume are connected to other conductive faults in such a way that a connected flow path is formed to allow CO_2 to leak to a compartment where there could be a potential impact. The connectivity of the conduits is related to the geometric characteristics of the system of conduits (i.e., distribution of conduits) between the storage reservoir and the compartment. For a site (which includes the storage formation and the geological formation above it) to be selected for GCS, some fault and fracture distribution data are expected to be available. However, the information on the conduit system is usually limited and highly uncertain. Therefore, it is a challenge to predict (1) if the conduits are connected, and if so, (2) the probability that a CO_2 plume will encounter the connected pathways. The objective of this work is to provide a framework for estimating the likelihood that CO_2 will intersect a conductive conduit network that allows leakage to occur. The amount of

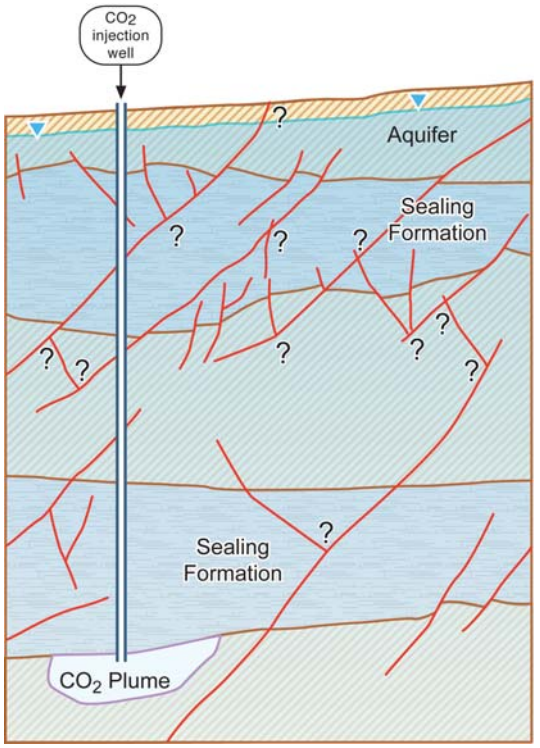

Fig. 1 Schematic geologic cross section (not to scale) showing CO_2 injection well, CO_2 plume, sealing formation, overlying formations, and potable ground water, along with conductive faults that may or may not intersect with each other as indicated by the question marks

leakage and the impact of leakage on the compartment are not within the scope of this study.

In the remainder of this paper, we develop the approach and demonstrate its applicability for a simplified geologic formation. The extension to a realistic CO_2 sequestration system is outlined in the conclusion section.

Approach

The proposed approach includes four steps: (1) estimate a critical value α_c for parameter α, which is related to the density of conduits (faults and fractures) in the conduit length distribution model, such that if $\alpha = \alpha_c$, there is a 50% probability that the system is connected between the storage formation and a compartment; (2) use numerical simulations to estimate the probability that the CO_2 plume will encounter the connected conduits for a system with α that is above or slightly less than α_c for various distributions of conduits, system size and CO_2 plume sizes; (3) construct fuzzy rules that relate information about the conduit system and CO_2 plume size to leakage probability based on simulation results from the previous step; and (4) using the fuzzy rules developed in the previous step, predict the probability that CO_2 will leak from the storage formation to a compartment through connected conduits for a given system. Each of these steps will be described in detail in the subsequent subsections.

One of the main advantages of the proposed approach is that conduit network generation (Step 2), which is the computationally intensive task, only needs to be done once. Once rules are generated (Step 3), P_{leak} can be easily determined in Step 4 for any given system with a known or estimated conduit distribution.

While the approach can be expanded to be applicable to realistic geologic formations, the four steps are illustrated for a simplified system with the following assumptions:

- The system is a square, two-dimensional (2D) cross section with sides of length L. The compartment where impacts occur is located at the top of the system and is of size L.
- Faults/fractures are represented by line segments in 2D.
- Faults/fractures are randomly positioned and oriented.
- Only conductive faults/fractures are considered.
- Faults/fractures follow a power-law length distribution.
- The details of the proposed approach are discussed next.

Calculation of critical α

In this part, we first justify and describe the power-law fault length distribution we choose for the subsequent analysis. Then we will introduce the percolation threshold, which is related to the connectivity of the fault system. Finally, we provide the relationship between the percolation threshold and the critical value of α (the coefficient in the power-law length distribution).

Distribution of fault length

Extensive studies have been done to characterize fault systems. The power-law distribution is the most widely used model to describe the fault-length distribution (Gudmundsson 1987; Scholz and Cowie 1990; Segall and Pollard 1983). Other statistical descriptions used to characterize fault length include the lognormal distribution (Priest and Hudson 1981; Rouleau and Gale 1985), the exponential distribution (Carbotte and Macdonald 1994; Cowie et al. 1993, 1995; Dershowitz and Einstein 1988), the stretched exponential distribution (Laherrere and Sornette 1998), and the Gamma distribution (Davy 1993; Kagan 1997; Main 1996; Sornette and Sornette 1999). A detailed review of fault characterization distributions can be found in Bonnet et al. (2001). Based on the arguments of Bonnet et al. (2001), we will use a power-law function to describe the fault-length distribution.

The power-law distribution of a fault system is given by:

$$n(l) = \alpha\, l^{-a} \tag{1}$$

where $n(l)dl$ is the number of faults having a length in the interval $[l, l + \mathrm{d}l]$, α is a coefficient of proportionality that reflects fault density and depends on the system size L (assuming a square system with sides of length L), and a is an exponent, which typically varies between one and three. It is apparent from Eq. 1 that the power-law distribution contains no characteristic length. This is the key argument for using power laws to describe fault growth processes (Bonnet et al. 2001).

Percolation threshold

Percolation theory (Stauffer and Aharony 1992) has been applied to study the connectivity of fault systems. In percolation theory, a percolation parameter p is used as an average measure of the geometrical properties, generally related to the density of faults, which also provides information on the connectivity of the system. For a 2D system (size L) with a total number of N faults of constant length l, the percolation parameter p is defined as:

$$p = Nl^2/L^2. \tag{2}$$

The percolation threshold p_c is defined as the critical p value below which the fault system is not connected (on average), whereas when p is above the critical value p_c, the system is on average connected. In other words, 50% of the systems at the percolation threshold are connected.

The percolation threshold p_c can be obtained from excluded area arguments (Balberg et al. 1984) or numerical simulation (Robinson 1983). For a power-law length distribution (Eq. 1) of a fault network, Bour and Davy (1997) demonstrated that the percolation threshold $p_c(L)$ does not vary significantly with L. For any value of a, the computed values of $p_c(L)$ are around 5.6 in two dimensions. Therefore, 5.6 will be used as the first approximation for p_c in the next step.

The purpose of the first step is to find a critical parameter α_c at percolation threshold, where α_c is related to the critical density of conduits at which the network is on average connected for a given system size.

Determining critical parameter α_c

Studies on the connectivity of faults include both faults with constant length (Balberg et al. 1984, 1991; Gueguen and Dienes 1989; Stauffer and Aharony 1992) and faults with a power-law distribution (Bour and Davy 1997, 1998; Renshaw 1999). Here, we will adopt some results from the study of Bour and Davy (1997) to find the critical density of a fault system.

Bour and Davy (1997) presented an analytical expression for the percolation threshold for a fault system following a power-law length distribution. The analysis is based on the relative contribution of small and large faults for defining the percolation threshold, e.g., the percolation threshold should be the sum of two terms describing the behavior of "small" and "large" faults:

$$p_c(L) = \int_{l_{\min}}^{L} \frac{n(l)l^2}{L^2}\mathrm{d}l + \int_{L}^{l_{\max}} n(l)\mathrm{d}l. \tag{3}$$

where L is the system size, $l_{\min}$ and $l_{\max}$ are the smallest and largest conduit lengths considered in the system, and $n(l)$ is the probability density function (pdf) of fault length distribution (in our case, the power-law distribution). If $l_{\max}$ is less than L, the second term on the right-hand side of Eq. 3 drops out and the first term integrates to $l_{\max}$ instead of L.

By inserting Eq. 1 for $n(l)$ into Eq. 3, and setting p_c to 5.6, an expression for the critical fault density α_c is obtained

$$\int_{l_{\min}}^{L} \frac{\alpha_c l^{2-a}}{L^2}\mathrm{d}l + \int_{L}^{l_{\max}} \alpha_c l^{-a}\mathrm{d}l = 5.6. \tag{4}$$

As shown in Figure (9) of Bour and Davy (1997), p_c varies very slightly around 5.6 for different values of a and L, with the largest discrepancy of $p_c \approx 7$. In this case, the calculated α_c using the above equations underestimates α_c by about 25%, which is conservative. We assume $p_c(L) = 5.6$ provides a good first approximation for a finite system with a power-law distribution of conduit length. Corrections for $p_c(L)$ due to finite-size effect and different values of the exponent a in the pdf of fault length distribution will not be made in this analytical formulation. However, the overestimation of α_c caused by these effects will be considered in the numerical simulation by performing simulations for systems with α less than α_c, until the resulting P_{leak} is considered small enough to be ignored.

Renshaw (1999) demonstrated that the connectivity of power-law length distribution networks is insensitive to the lower cutoff length $l_{\min}$ as long as this length is sufficiently small. However, it is sensitive to the higher cutoff length $l_{\max}$. This effect will also be considered in the numerical simulation.

The length scales in both Eqs. 1 and 4 can be normalized by the smallest fault size $l_{\min}$: $l_s = l/l_{\min}$, $L_s = L/l_{\min}$, and $l_{\max\ s} = l_{\max}/l_{\min}$. By defining $\alpha_s = \alpha\, l_{\min}^{-a+1}$, we obtain:

$$n(l)\,\mathrm{d}l = \alpha l^{-a}\mathrm{d}l = \alpha l_{\min}^{-a+1}\frac{l^{-a}}{l_{\min}^{-a}}d\frac{l}{l_{\min}} = \alpha_s l_s^{-a}\mathrm{d}l_s. \tag{5}$$

For $a \neq 3$, and if $L \geq l_{\max}$,

$$\begin{aligned} p_c(L) &= \int_{l_{\min}}^{l_{\max}} \frac{n(l)l^2}{L^2}\mathrm{d}l = \int_{1}^{l_{\max s}} \frac{\alpha_{sc} l_s^{2-a}}{L_s^2}\mathrm{d}l_s \\ &= \frac{\alpha_{sc}}{L_s^2}\left[\frac{1}{a-3} - \frac{l_{\max s}^{3-a}}{a-3}\right] = 5.6. \end{aligned} \tag{6}$$

Where α_{sc} is the critical value of α_s when the system is at the percolation threshold p_c whereas if $L < l_{\max}$,

$$\begin{aligned} p_c(L) &= \int_{1}^{L_s} \frac{\alpha_{sc} l_s^{2-a}}{L_s^2}\mathrm{d}l_s + \int_{L_s}^{l_{\max s}} \alpha_{sc} l_s^{-a}\mathrm{d}l_s \\ &= \alpha_{sc}\left[\frac{1}{a-3}\frac{1}{L_s^2} + \left(\frac{1}{3-a} + \frac{1}{a-1}\right)L_s^{1-a} + \frac{1}{1-a}l_{\max s}^{1-a}\right] \\ &= 5.6. \end{aligned} \tag{7}$$

For $a = 3$, and if $L \geq l_{\max}$

$$p_c(L) = \alpha_{sc}\frac{1}{L_s^2}\ln l_{\max s} = 5.6 \tag{8}$$

whereas if $L < l_{\max}$,

$$p_c(L) = \alpha_{sc}\left[\frac{1}{L_s^2}\ln L_s + \frac{1}{a-1}L_s^{-2} + \frac{1}{1-a}l_{\max s}^{1-a}\right] = 5.6. \tag{9}$$

The expression for α_{sc} shows that a larger system size L_s corresponds to a larger α_{sc}; a larger $l_{\max\ s}$ corresponds to a smaller α_{sc}; and a larger exponent a (larger portion of smaller faults) corresponds to a larger α_{sc}.

For a given system, we can calculate the critical parameter α_{sc} and compare it to the actual parameter α_s. If the actual density is much smaller than the critical value, we can conclude that the system is not connected and the CO_2 plume will not be able to leak out through the fault system. For systems with α_s around or above its critical value, the steps described in "Conduit network generation to determine P_{leak}" and "Construct fuzzy rules for calculating P_{leak}" need to be performed.

The above formulation works for a square system and for conduits with random orientations. For an anisotropic system—a system with different horizontal and vertical connectivity on average—the percolation theory still holds, but with a modified expression for the percolation threshold (Masihi et al. 2006).

Conduit network generation to determine P_{leak}

The purpose of this step is to provide a basis to form fuzzy rules for P_{leak} for systems of different size, fault geometries, and CO_2 plume size through numerical generation of fault networks using a limited number of parameters. We assume that vulnerable resources are located at the top of the system. This means the probability that a connected pathway (connected also to the top of the system), encounters a compartment is 1. In this case, P_{leak} depends on only two unknowns, i.e., whether the system is connected (U1), and whether a connected pathway intersects the CO_2 plume (U2). U2 depends on both the number of connections and the size of the CO_2 plume.

Two types of uncertainties are considered. The first stems from our lack of knowledge of the system, specifically, parameters used to describe fault-length distribution and reservoir properties used to estimate CO_2 plume size. This uncertainty will be considered by using fuzzy-rule-based modeling to propagate the uncertainty of the input parameters to the estimation of P_{leak}. We vary system size and fault-length distribution parameters to generate fault networks and estimate P_{leak} for various CO_2 plume sizes. The second type of uncertainty is the uncertainty in the generation of the discrete fault network itself. Even for systems with the same parameters (e.g., system size and fault distribution), the generated network could have very different connectivities. To consider this uncertainty on the evaluation of the leakage probability P_{leak}, multiple realizations of random discrete fault networks using the same parameter set are generated. The average and 95% confidence interval are used to interpret the results.

The parameters varied in the fault network generation and P_{leak} calculations are the normalized system size L_s, the normalized maximum fault length $l_{max\ s}$, the exponent a, the ratio of $r = \alpha_s/\alpha_{sc}$ (representing the system's actual fault density compared to that at the percolation threshold), and the normalized plume size M_s, which is the CO_2 plume size divided by the smallest fault size l_{min}.

The total number of faults that exist in the system is obtained by integrating the left side of Eq. 1:

$$N = \int_1^{l_{max\,s}} \alpha_s l_s^{-a} \mathrm{d}l_s = \frac{\alpha_s}{1-a}\left(l_{max\,s}^{1-a} - 1\right). \tag{10}$$

To generate a fault network, three parameters for characterizing the geometry of individual faults need to be determined, namely location, orientation, and length. In our simulation, we locate the center of conduits randomly in the system. Fault orientation is also random, e.g., uniformly sampled from all directions. Fault length (l) is sampled from its power-law distribution. Note that hydraulic properties of the faults are not needed in this study, as no leakage flux is determined, and because fault network connectivity is estimated using geometrical parameters of the hydraulically conductive faults.

Once the fault network is generated using a given set of L_s, $l_{max\ s}$, a, and $r = \alpha_s/\alpha_{sc}$, to calculate P_{leak}, we need to (1) remove the unconnected faults (starting from the top of the system) and examine if a connected pathway can be established when the bottom of the system is reached, and (2) calculate the probability (P_{inter}) that the CO_2 plume encounters this/these connected pathway(s). The first step will be explained in detail in the illustrative example. In the second step, for systems where no connected pathway is established, $P_{inter} = 0$; for systems where a connected pathway is found, an "effective connection" concept and a moving average method are used to calculate P_{inter}. All connections encountered within M_s/L_s (i.e., the normalized CO_2 plume size divided by the normalized system size) are counted as a single effective connection. With each of the assumed plume sizes, P_{inter} is obtained by averaging effective connections for the CO_2 plume at different locations. In other words, a moving average is performed by moving the CO_2 plume along the caprock and checking if it encounters a connected pathway. If it does, we assign a number of one, and if it does not, we assign a number of zero. The final averaged number is P_{inter}—the probability that the CO_2 plume encounters a conductive fault (or fault zone) that is connected to other conductive conduits and serves as a pathway for CO_2 to escape from the reservoir and migrate to compartments. As discussed earlier, P_{leak} depends not only on P_{inter}, but also on the probability (P_{con}) that a connected pathway exists in the system. To consider P_{con}, for each parameter set, we average P_{inter} from all realizations (including the ones with $P_{inter} = 0$ and the ones obtained using moving average method) to obtain P_{leak}. In this way, P_{con} is implicitly considered in the P_{leak} calculation.

As a result of this procedure, the following statement can be made for each parameter set:

IF $L_s = L_1$ **AND** $l_{\max s} = l_1$ **AND** $a = a_1$ **AND** $r = r_1$ **AND** $M_s = M_1$,

THEN P_{leak}, is b.

Here, L_1, l_1, a_1, r_1 ($r_1 \geq 1$), and M_1 are the numerical values of the varying parameters in the simulation, covering all likely values. b is the calculated P_{leak} for each parameter set. Up to this point, both input variables and output variable are crisp numbers rather than fuzzy numbers.

Construct fuzzy rules for calculating P_{leak}

Fuzzy logic is viewed as a system of concepts, principles, and methods for dealing with models of reasoning that are approximate rather than exact (Novak and Perfilieva 2000). Due to its strength in dealing with uncertainty, ambiguity, and imprecision that one often encounters in modeling natural systems, fuzzy logic has been successfully applied to earth sciences including areas of surface and subsurface hydrology (e.g., Bardossy 1996; Bardossy et al. 2005; Bardossy and Disse 1993; Dou et al. 1999; Hundecha et al. 2001), water resources management and risk assessment (e.g., Kumar et al. 2006; Panigrahi and Mujumdar 2000; Shrestha et al. 1996; Uricchio et al. 2004), and soil science (e.g., Bardossy and Lehmann 1998; Mays et al. 1997; McBratney and Degruijter 1992; McBratney and Odeh 1997; Odeh et al. 1992).

Fuzzy-rule based modeling represents a complex system with imprecise, vague, and uncertain information by means of fuzzy rules. A fuzzy rule (the ith rule) consists of a set of arguments $C_{i,k}$ (kth argument, $k = 1,\ldots, K$) in the form of fuzzy sets with membership functions $\mu_{C_{i,k}}$, and a consequence B_i, which is also in the form of a fuzzy set with a membership function μ_{B_i}. The membership function $\mu_{C_{i,k}}$ expresses the grade or degree of membership of element x in $C_{i,k}$. A simple fuzzy rule statement reads as follows:

IF x_1 is $C_{i,1}$ **AND** x_2 is $C_{i,2}$ **AND** ... **AND** x_K is $C_{i,K}$,
THEN B_i.

The use of fuzzy sets (instead of crisp numbers) in the fuzzy statements allows the rules to be used conveniently for both descriptive and quantitative purposes. In addition, these fuzzy rules can be partially or simultaneously fulfilled. This means a rule can have partial applicability, or it is possible to have a few partially applicable rules combined.

The degree of fulfillment (DOF) v is used to quantify the truth grade corresponding to the fulfillment of the conditions of a rule (the ith rule) for given premises ($x_1,\ldots, x_K$).

$$v(x_1 \textbf{ AND } x_2 \textbf{ AND} \ldots \textbf{AND } x_K) = \mu_{C_{i,1}}(x_1)\mu_{C_{i,2}} \times (x_2) \cdots \mu_{C_{i,K}}(x_K). \quad (11)$$

Fuzzy rules can be developed using expert opinions, existing data, and qualitative information. Alternatively, fuzzy rules can be generated through numerical simulations (Bardossy and Disse 1993; Bardossy and Duckstein 1995; Dou et al. 1999). In our case, we use results from Step 2 as the training set to construct fuzzy rules. Since the training set is from numerical simulations covering the feasible input parameter (fault network parameters and plume size), we know the input data structure (e.g., exponent a between 1 and 3, we can define a few fuzzy numbers evenly distributed between 1 and 3). We use the weighted counting algorithm proposed by Bardossy and Duckstein (1995) to construct fuzzy rules. In this method, the rule premises are defined explicitly, and responses to the rule are defined using all simulation data sets. The training set Γ is written as:

$$\Gamma = \{(x_1(s), \ldots, x_K(s), b(s)); \quad s = 1, \ldots, S\} \quad (12)$$

where b refers to the consequence, in our case it is P_{leak}. And S is the total number of data (in our case, the number of parameter set in the simulation).

The algorithm can be described as follows:

Define the membership functions of the premises. There are five arguments ($K = 5$) in our case: system size L_s, largest fault size $l_{\max\ s}$, exponent a, ratio $r = \alpha_s/\alpha_{sc}$, and ratio M_s/L_s (for convenience, we use the ratio of plume size over system size to represent relative plume size). If triangular membership functions are used for $C_{i,k}$ (the kth argument of the ith rule), define the fuzzy number ($c^-_{i,k}$, $c^1_{i,k}$, $c^+_{i,k}$) as shown in Fig. 2.

Calculate the DOF $\boldsymbol{v}_i$ of each rule for each premise vector [$x_1(s)$,…, $x_k(s)$] corresponding to the training set.

Select a number $\varepsilon > 0$ such that only responses with a DOF of at least ε will be considered in the construction of the rule response. The corresponding response is assumed to be a triangular fuzzy number (β_i^-, β_i^1, β_i^+), where

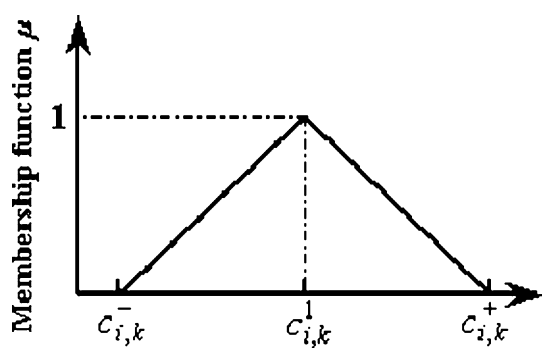

Fig. 2 Triangular membership functions for a fuzzy number

$$\beta_i^- = \min_{v_i(s) > \varepsilon} b(s) \tag{13}$$

$$\beta_i^1 = \frac{\sum_{v_i(s) > \varepsilon} v_i(s) b(s)}{\sum_{v_i(s) > \varepsilon} v_i(s)} \tag{14}$$

$$\beta_i^+ = \max_{v_i(s) > \varepsilon} b(s). \tag{15}$$

The resulting fuzzy rules have the following format:

IF $L_s = (L_{si}^-, L_{si}^1, L_{si}^+)$ **AND** $l_{\max s} = (l_{\max\, s\, i}^-, l_{\max\, s\, i}^1, l_{\max\, s\, i}^+)$ **AND** $a = (a_i^-, a_i^1, a_i^+)$ **AND** $\alpha_s/\alpha_{sc} = (r_i^-, r_i^1, r_i^+)$ **AND** $M_s/L_s = (m_i^-, m_i^1, m_i^+)$, **THEN** the probability that a CO_2 plume escapes through a connected conduit system is $P_{\text{leak}} = (\beta_i^-, \beta_i^1, \beta_i^+)$.

Calculate P_{leak} for a given system

For a given system, the first step is to calculate α_{sc} and to compare it to α_s. If the two are the same, P_{leak} should be about 0.5. If the latter is much smaller, $P_{\text{leak}} = 0$. We will address how much smaller α_s has to be so P_{leak} can be ignored in the example problem. For systems with an α_s around or above α_{sc}, the above fuzzy rules are used to infer P_{leak}. There two commonly used inference models, referred to as Mamdani-type models (Mamdani and Assilian 1975) and the Takagi–Sugeno-type models (Takagi and Sugeno 1985). However, the Takagi–Sugeno-type system uses a single spike as the output membership function, where the output (consequence) is calculated as the weighted average of a few data points rather than integrating over the domain of the output fuzzy set. The method does not provide the uncertainty range of the output, which is what we need for our prediction result. The Mamdani-type inference system uses a maximum combination method to aggregate fuzzy rules. The method tolerates disagreement between rules, but it does not increase the membership function of the response if two rules give the same results. Furthermore, it does not give higher weights to rules with crisper answers, therefore it could make the response very vague.

We use the normalized sum combination method proposed by Bardossy and Duckstein (1995) to aggregate fuzzy rules. This method has the advantage of assigning more weight to rules with crisper answers than fuzzy answers (e.g., less weight on less certain answers). It is calculated as:

$$\mu_B(x) = \frac{\sum_{i=1}^{I} v_i \tau_i \mu_{B_i(x)}}{\max_u \sum_{i=1}^{I} v_i \tau_i \mu_{B_i(u)}} \tag{16}$$

where

$$\frac{1}{\tau_i} = \int_{-\infty}^{+\infty} \mu_{B_i}(x) \mathrm{d}x. \tag{17}$$

The division by the maximum of the summation is to ensure the resulting membership function is not >1. Moreover, if the centroid is used as the defuzzification method, the fuzzy mean can be simply calculated as:

$$M(B) = \frac{\sum_{i=1}^{I} v_i M(B_i)}{\sum_{i=1}^{I} v_i}. \tag{18}$$

Illustrative example

In the following example, we establish rules and predict P_{leak} for systems with a normalized system size L_s between 50 and 200, and a normalized largest fault size $l_{\max\ s}$ between 50 and 200. The exponent a in Eq. 1 is only considered for values between 1.1 and 3. The first step is to use Eqs. 6–9 to calculate the critical parameter α_{sc} and to compare it to the actual value α_s. Only systems around the percolation threshold (including a little less) or above are considered to be possible to have connected pathways.

Next, we determine how many fault networks with different sets of L_s, $l_{\max\ s}$, a, and α_s/α_{sc} are needed to construct robust fuzzy rules. We sample these parameters uniformly over the admissible range, making sure that no excessive extrapolations are needed when subsequently generating fuzzy rules. The admissible range for each of the parameters is listed in Table 1. The total number of parameter sets evaluated to generate fault networks is about 1,800. For each parameter combination, 100 discrete fault networks are generated, and for each realization, different plume sizes are considered to calculate P_{inter}. Then the leakage probability P_{leak} is obtained by averaging P_{inter} (for the same plume size) from the 100 realizations.

We use a fracture network generation code modified based on the one that was used by Liu et al. (2002) and Zhang et al. (2004). Figure 3 shows the power-law fault-length distribution (solid line) that is used to generate fault networks, and the actual generated fault-length distribution (symbols) for a system with $a = 1.5$, $L_s = l_{\max\ s} = 200$, and $r = 1$. As expected, the double-log plot shows a good fit for small faults with high frequency. For big faults with low frequency, since the number of faults can only be an integer, which is discrete, the samples are spread out around the analytical length distribution. The fault length distribution generated in the network is considered to coincide well with the specified power-law distribution. Figure 4 is an example of a discrete fault network for a system with $L_s = 100$, $l_{\max\ s} = 200$, $a = 1.1$, and $r = 1.1$. Figure 4a shows all generated faults. Starting with the

Table 1 Parameters used in fault network generation

a	L_s	$l_{\max\ s}$	α_s/α_{sc}	M_s
1.1–3.0	50–200	50–200	0.75–2.5	10-L_s

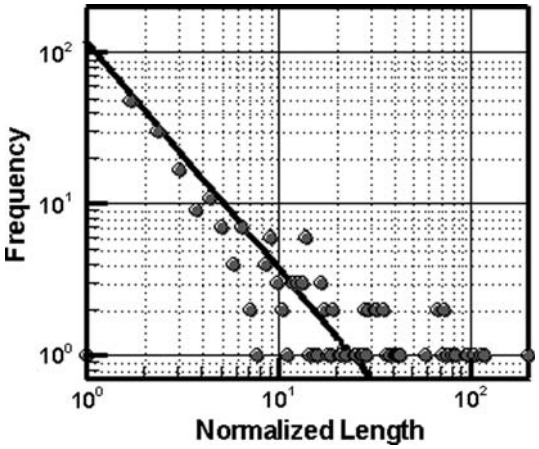

Fig. 3 Power law fault length distribution (*solid line*) and sample length distribution (*symbols*) for a system with $a = 1.5$, $L_s = l_{\max\ s} = 200$, and α_s/α_{sc}

faults intersecting the top of the system, we gradually find and plot only the faults that are connected (i.e., the unconnected faults are removed); the resulting network of connected faults is shown in Fig. 4b. The removal of unconnected faults is started at the top, so that the calculation of effective connections, which are identified at the bottom of the system, where CO_2 resides, remains zero until connectivity across the entire system is achieved. Based on this plot, we can easily find P_{leak} using the moving average method. Note that while there are three conduits connected to the CO_2 storage formation at the bottom of the system, two of the three connections are very close to each other. Such a cluster does not increase the leakage probability significantly compared to a single connection at that location. The simple moving average method described in "Conduit network generation to determine P_{leak}" properly accounts for this effect. For a normalized plume size $M_s = 10$, and a moving unit of $M_s/10 = 1$, we obtain $P_{\text{leak}} = 0.21$, which is smaller than the probability one might expect based on the total number of connections alone, but slightly larger than if there were only two, clearly separated faults.

Figure 5 shows a fault network with $a = 3.0$ and $r = 1.25$ (L_s and $l_{\max\ s}$ are the same as those used to generate Fig. 4). With increasing a, a larger portion of the total fault population consists of small faults. Consequently, the total number of faults and the critical value α_{cs} also increases. In this example, despite the very densely distributed small faults in the domain, the system is not connected (see Fig. 5b). Figure 6 shows a network generated using exactly the same system parameters as the network shown in Fig. 5, but for this realization the system is connected. Figures 4, 5 and 6 demonstrate that with increasing a, the portion of small faults in the entire

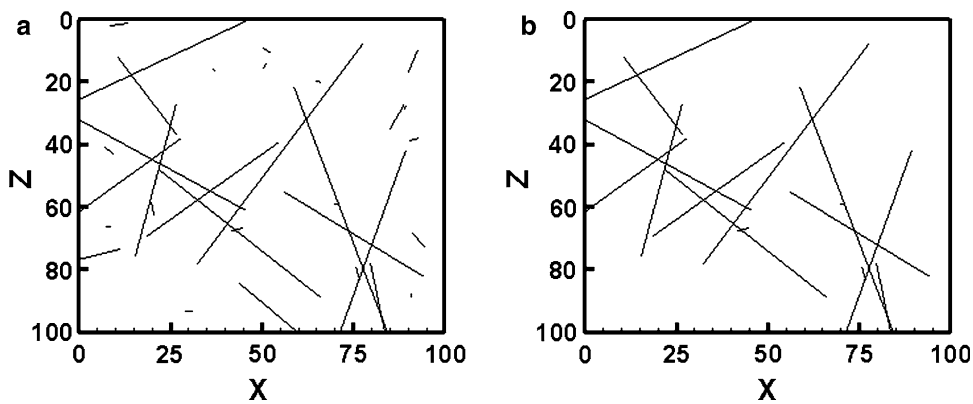

Fig. 4 A discrete fault network for a system with $a = 1.1$, $L_s = 100$, $l_{\max\ s} = 200$, $\alpha_s/\alpha_{sc} = 1.1$ (**a**) with all fractures and (**b**) with unconnected faults sequentially removed from the *top* to the *bottom*

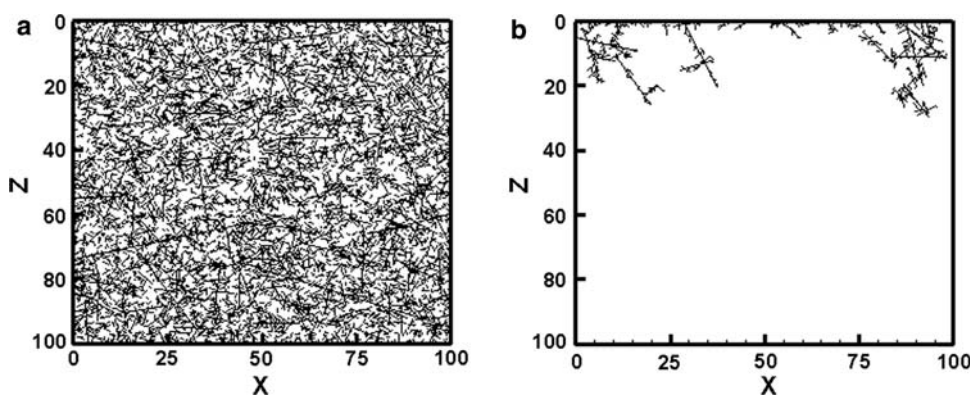

Fig. 5 A discrete fault network for a system with $a = 3.0$, $L_s = 100$, $l_{\max\ s} = 200$, $\alpha_s/\alpha_{sc} = 1.25$ (**a**) with all fractures and (**b**) with unconnected faults sequentially removed from the *top* to the *bottom*

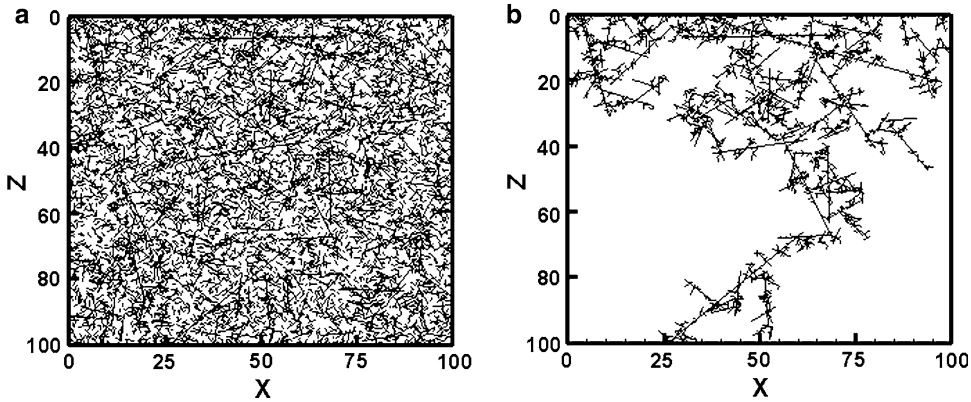

Fig. 6 A discrete fault network for a system with $a = 3.0$, $L_s = 100$, $l_{\max\ s} = 200$, $\alpha_s/\alpha_{sc} = 1.25$ (**a**) with all fractures and (**b**) with unconnected faults sequentially removed from the *top* to the *bottom*

distribution increases significantly, gradually making overall connectivity (should it exist) dominated by small rather than large faults.

In Fig. 7, we plot P_{leak} and its 95% confidence interval as a function of CO_2 plume size for a system with $a = 1.5$, $l_{\max\ s} = 100$, $L_s = 100$, and different ratios $r = \alpha_s/\alpha_{sc}$. The value of P_{leak} at $M_s = L_s$ (plume size equals system size) is equal to the probability that the system is connected. If the theoretical percolation threshold were accurate, P_{leak} is expected to be around 0.5 for $r = 1$. As it turns out, this is not always the case, as demonstrated in Fig. 7 where P_{leak} is only about 0.25. There are two reasons for this. First, the theoretically derived percolation threshold of 5.6 used in the calculation needs to be corrected to account for differences between the systems studied here and that used to develop the theory. Specifically, there are differences in (a) the exponent a, (b) the system size, leading to a finite-size effect, and (c) the cut-off fault size for the largest fault. Secondly, we generated the faults by randomly locating the centers of the fault within the system domain. Because no faults are generated if their center points are outside the model domain, the fault density near the domain edge is somewhat reduced. Therefore, the estimated P_{leak} ($r = 1$, $M_s = L_s$) tends to be smaller than 0.5. The second reason is an artifact of the fault network generation, which could be eliminated by generating additional faults with the center points outside the model domain. Such a revision, however, is only justified if geological evidence suggests that fault density indeed does not show an edge effect near lithological or tectonic boundaries.

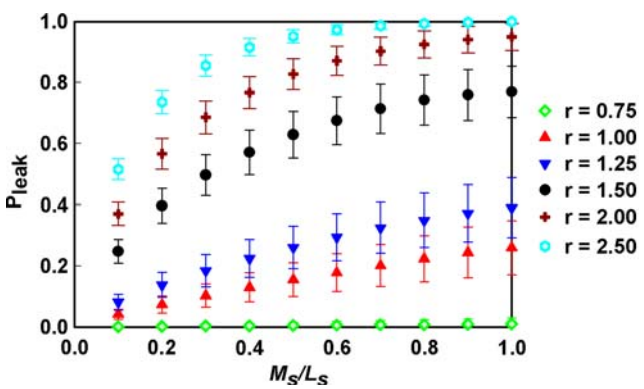

Fig. 7 P_{leak} as a function of CO_2 plume size for a system with $a = 1.5$, $l_{\max\ s} = 100$, $L_s = 100$, and different values $r = \alpha_s/\alpha_{sc}$. The error bars indicate 95% confidence intervals

The original values we considered for r are in the range between 1 and 2.5. However, theoretically P_{leak} ($r = 1$) at percolation threshold should be about 0.5, which is not small enough for us to ignore the likelihood of CO_2 leakage. This is also demonstrated by simulation results (e.g., Fig. 7). We gradually add new simulations with r smaller than 1 to the set of fault networks. These additional simulation results indicate that if $r = 0.75$, P_{leak} at ($r = 0.75$, $M_s = L_s$) is smaller than 0.05. If this is considered to be acceptably small, we change the lower bound for generating fault network, also for generating fuzzy rules to $r = 0.75$. If $r = 2.5$, we consider P_{leak} to be high enough for us to conclude the likelihood of CO_2 leakage to compartments is significant and close to certain. Thus, this value is used as the upper bound for fuzzy rule generation.

As we mentioned in the previous section, Bour and Davy (1997) showed in their Fig. 9 that p_c varies between 5 and 7 for systems that have a size between 10 and 100, with exponent a between 1.8 and 3.2. This means that the needed correction for α_{sc} could be 25% or even higher. To confirm this assumption, we show results for $r = 1.25$ in Fig. 7, the expected connectivity is approximately 0.4, with the 95% confidence interval between 0.3 and 0.5. This supports the validity of the assumption.

The 95% interval appears to be bigger for r between 1 and 2, because for fault networks near the percolation threshold, the chance that a system is or is not connected depends on the random presence or absence of a few critical faults and is thus highly random.

Table 2 is a list of α_{sc} value calculated using Eqs. 6–9, as a function of $l_{\max\ s}$ for both $a = 1.5$ and $a = 3.0$, $L_s = 100$. When $l_{\max\ s}$ increases from 50 to 100, α_{sc} is

Table 2 α_{sc} values for different a and $l_{max\ s}$

$l_{\max\ s}$	50	100	200
$a = 1.5$	238	84	45
$a = 3.0$	14,314	12,160	10,969

reduced by about 65% for $a = 1.5$, because large faults dominate the connectivity; on the other hand, for $a = 3.0$, small faults dominate the connectivity, and therefore the reduction of α_{sc} is much less, only 15%.

For each parameter set, 100 realizations with different seed numbers are created to consider the uncertainty in fault network generation. To construct fuzzy rules, the averaged results from these 100 realizations are used. Now we have a database with a total of S ($\sim$21,000) sets: $\{(x_1(s), x_2(s), x_3(s), x_4(s), x_5(s), b(s)); \quad s = 1, \ldots S\}$, with $x_1(s)$ representing exponent a, $x_2(s)$ representing normalized system size, $x_3(s)$ representing normalized largest fault size, $x_4(s)$ representing ratio $r = \alpha_s/\alpha_{sc}$, $x_5(s)$ representing $m = M_s/L_s$, and $b(s)$ representing P_{leak}. The first step in generating fuzzy rules is to define rule structure. For each argument, we define its fuzzy numbers as listed in Table 3.

With the structure listed in Table 3, there will be a total of 1,620 rules (a combination of 5 arguments: $5 \times 3 \times 3 \times 6 \times 6$). For each parameter set, the DOF of each rule is calculated. The ones that have a DOF value larger than $\varepsilon = 0.2$ are kept for calculating the consequences using Eqs. 13–15.

To demonstrate how to predict P_{leak} for a given system with uncertainty, we use a simple example that only considers one premise. We consider a system with the exponent $a = 1.5$, both normalized system size and largest fault are 100, and the normalized plume size is 40 (e.g., 40% of the system size). The ratio $r = \alpha_s/\alpha_{sc}$ is considered to be uncertainty, estimated to be greater than but close to 1, somewhere between 1 and 1.25.

Corresponding rules (with the same a, L_s, $L_{\max\ s}$, and M_s) are used to predict P_{leak}. If we assume r can be represented by two fuzzy numbers as shown in Fig. 8, with a membership function of 0.5 for both, the two rules will have a DOF of 0.5, while all the other rules will have a DOF of zero. These two rules are applied to estimate P_{leak}:

Rule 1:

IF $a = (1.1, 1.5, 2.0)$ AND $L_s = (50, 100, 200)$ AND $l_{\max\ s} = (50, 100, 200)$
AND $r = (0.75, 1.0, 1.25)$ AND $r_{\text{p}} = (0.2, 0.4, 0.6)$
THEN $P_{\text{leak}} = (0.01, 0.12, 0.18)$

Rule 2:

IF $a = (1.1, 1.5, 2.0)$ AND $L_{\text{s}} = (50, 100, 200)$ AND $l_{\max\ s} = (50, 100, 200)$
AND $r = (1.0, 1.25, 1.5)$ AND $r_{\text{p}} = (0.2, 0.4, 0.6)$
THEN $P_{\text{leak}} = (0.1, 0.3, 0.55)$

The individual responses from the two rules are shown as thin black lines in Fig. 9. The area under the membership function curve is a measure of how uncertain the estimated P_{leak} is: the larger the area, the more uncertain it is. In this case, the area for rule 1 is 0.085 and that for rule 2 is 0.27. Rule 1 has a crisper result than Rule 2; therefore, it has a larger weight (the weight is inverse to the area) in the final combined results. The final P_{leak} and its membership function are shown as thick red lines. The defuzzified P_{leak} value (using the centroid method) is 0.21.

We apply the fuzzy rules to the same system as shown in Fig. 7. The predicted P_{leak} using fuzzy rules are also fuzzy numbers. After defuzzification, we plot them in Fig. 10. Although the results are consistent with results from the Monte Carlo simulation, they are not identical because by using fuzzy-rule based modeling, we have considered the uncertainty in the input parameters, whereas the multiple realizations for the same parameter set only considered the randomness of equally probable fault networks. When we talk about r being approximately 1.5, we have implicitly include the possibilities that r could be larger or smaller than 1.5. As a result, the estimated P_{leak} contained uncertainty in r. Since the combinations of the rules are weighted by both the DOF of a rule and the area of the membership function of the rule outcome, the defuzzified P_{leak} could be smaller or larger than the Monte Carlo simulation, which did not consider the input parameter uncertainty. However,

Table 3 Fuzzy numbers used in the premises

a	L_s	$l_{\max\ s}$	α_s/α_{sc}	m
(1.1, 1.1, 1.5)	(50, 50, 100)	(50, 50, 100)	(0.75, 0.75, 1.0)	(0.0, 0.0, 0.2)
(1.1, 1.5, 2.0)	(50, 100, 200)	(50, 100, 200)	(0.75, 1.0, 1.25)	(0.0, 0.2, 0.4)
(1.5, 2.0, 2.5)	(100, 200, 200)	(100, 200, 200)	(1.0, 1.25, 1.5)	(0.2, 0.4, 0.6)
(2.0, 2.5, 3.0)			(1.25, 1.5, 2.0)	(0.4, 0.6, 0.8)
(2.5, 2.5, 3.0)			(1.5, 2.0, 2.5)	(0.6, 0.8, 1.0)
			(2.0, 2.5, 2.5)	(0.8, 1.0, 1.0)

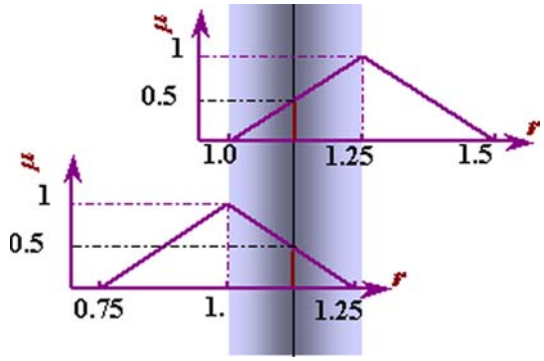

Fig. 8 The membership functions for the two fuzzy numbers that are used to represent "r is greater but close to 1"

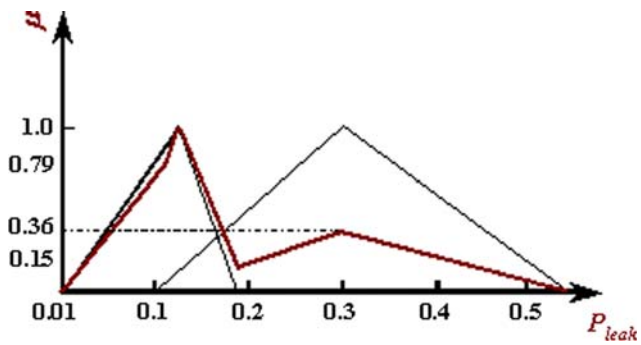

Fig. 9 Membership functions of P_{leak} (*thick red line*) for a system ($a = 1.5$, $L_s = l_{\max\ s} = 100$, $M_s = 40$) with "r greater but close to 1". The two individual rule consequences are shown as thin *black lines*

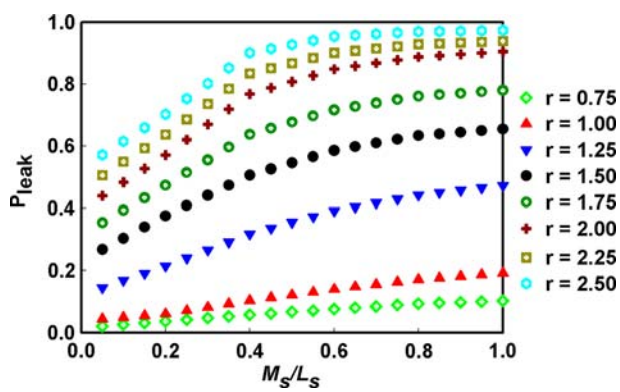

Fig. 10 Fuzzy-rule based prediction of P_{leak} as a function of r_p (CO_2 plume size divided by system size) for a system with $a = 1.5$, $l_{\max\ s} = 100$, $L_s = 100$, and different values of $r = \alpha_s/\alpha_{sc}$

there are two exceptions in our prediction. When $r = 0.75$, the predicted P_{leak} values using fuzzy rules are always larger than the Monte Carlo simulation results, and when $r = 2.5$, the predicted P_{leak} values are always smaller than the Monte Carlo simulation results (see Figs. 7, 10). This is because we use a triangular fuzzy number (0.75, 0.75, 1.0) for the statement "r is likely to be 0.75 or slightly above." The uncertainty in this number means a possibility that r is higher than 0.75, but not lower. This effect makes P_{leak} at $r \approx 0.75$ always higher than P_{leak} at $r = 0.75$. Similarly, because we define a triangular membership function (2.0, 2.5, 2.5) for the statement "r is close to 2.5 or slightly smaller", the uncertainty in r means a possibility that r is smaller than 2.5, but not larger, and the P_{leak} values predicted using fuzzy rules are always smaller.

Conclusions and practical considerations

In this paper we presented a method to estimate a limiting factor controlling the probability of CO_2 leakage through a fault or fracture system, namely the probability (P_{leak}) of the plume intersecting a connected network of faults or fractures that also intersects a vulnerable resource. The proposed method includes (1) the estimation of the connectivity of the fault system using percolation theory; (2) the estimation (for a limited number of systems) of the probability that CO_2 plumes with different sizes encounter a connected system of conduits; (3) the construction of fuzzy rules for calculating P_{leak} considering uncertainty in the input parameters; and (4) predictive estimation of P_{leak} for a given system. The method was designed to fit in the CF, where the risk associated with CO_2 leakage is the product of the leakage likelihood and the impact of that leakage event. The rules that are generated from this study will be stored within the CF model. When a site needs to be evaluated as a potential geological CO_2 sequestration site, step (4) needs to be performed, and a leakage likelihood will be passed on for the final calculation of risk.

The study was done for a two-dimensional system. Representing an inherently three-dimensional system with a two-dimensional model yields conservative leakage probability estimates, because the third dimension is implicitly assumed to be connected. Nevertheless, the concept and methods described in this study can also be applied to more realistic three-dimensional fault-systems. Analyzing fault systems in three dimensions requires modifying (1) the percolation threshold value, (2) the expression of the percolation threshold (Bour and Davy 1998), (3) generating three-dimensional fault networks, (4) evaluating leakage probability using realistic plume configurations, and (5) recreating fuzzy rules.

The assumptions for the approach include a square system and randomly oriented conduits. However, for a real site, the system is unlikely to be square, and faults have preferential orientations. Both situations lead to a preferential connection in one direction (the short direction), and less likely connectivity in the other direction. This effect is

referred to as anisotropy, which can be accounted for following the method proposed by Masihi et al. (2007).

The main computational effort resides in the numerical generation of the fault networks and finding the connected pathway(s). In our case, we generated about 1,800 × 100 (for each system parameter there are 100 realizations) fault networks. The varying of plume size is not included in this number since once the connected pathway is found, there is not much computational effort involved in calculating the P_{inter} for various plume sizes. However, fault network generation only needs to be done once to provide the basis for constructing the fuzzy rules; predictive simulations are then performed very efficiently using these fuzzy rules. After we include the plume size in the input parameter set and average results of the 100 realizations, we have about 21,000 datasets to generate rules. If needed, additional networks can be added to the database to extend the input parameter ranges. By using fuzzy-rule based modeling, we can predict P_{leak} for systems that have characteristics different from the ones we have in the database (obtained from fault network generation), as well as the uncertainty of P_{leak}, by propagating the uncertainty in the input parameters. The randomness in the fault network generation is considered by generating multiple realizations for the same system.

Brine leakage through a fault or fracture system from the reservoir may also lead to environmental impact. Although we focus the application of the proposed method to evaluate the probability of CO_2 leakage, the approach can also be applied to estimate brine leakage probability. An effective brine plume size, analogous to the CO_2 plume size, could be defined as the size of the pressure perturbation above some cut-off value. Using the proposed method, a relationship can be established between the leakage probability and the region of pressure perturbation.

If GCS becomes a viable mitigation option to address CO_2 emissions, a large number of sites with different amounts of characterization data and degrees of uncertainty will need to be evaluated. The proposed method provides a tool for a preliminary evaluation of leakage likelihood through fault systems. In the future, this simplified model will be refined to better represent the dimensionality and fault distributions at actual sites under evaluation.

Acknowledgments This work was supported in part by the CO_2 Capture Project (CCP) of the Joint Industry Program (JIP), and by Lawrence Berkeley National Laboratory under Department of Energy Contract No. DE-AC02-05CH11231. We thank Keni Zhang for providing his fracture network generation code. We also thank Christine Doughty and Hui-Hai Liu (LBNL) for constructive reviews, and Scott Imbus (Chevron) and Cal Cooper (ConocoPhillips) for support and encouragement.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Balberg I, Anderson CH, Alexander S, Wagner N (1984) Excluded volume and its relation to the onset of percolation. Phys Rev B 30(7):3933–3943

Balberg I, Berkowitz B, Drachsler GE (1991) Application of a percolation model to flow in fractured hard rocks. J Geophys Res Solid Earth Planets 96(B6):10015–10021

Bardossy A, Duckstein L (1995) Fuzzy rule-based modeling with applications to geophysical, biological and engineering systems, p 110. CRC Press, Boca Raton

Bardossy A (1996) The use of fuzzy rules for the description of elements of the hydrological cycle. Ecol Modell 85(1):59–65

Bardossy A, Disse M (1993) Fuzzy rule-based models for infiltration. Water Resour Res 29(2):373–382

Bardossy A, Lehmann W (1998) Spatial distribution of soil moisture in a small catchment. Part 1: Geostatistical analysis. J Hydrol 206(1–2):1–15

Bardossy A, Bogardi I, Matyasovszky I (2005) Fuzzy rule-based downscaling of precipitation. Theor Appl Climatol 82(1–2):119–129

Bonnet E, Bour O, Odling NE, Davy P, Main I, Cowie P, Berkowitz B (2001) Scaling of fracture systems in geological media. Rev Geophys 39(3):347–383

Bour O, Davy P (1997) Connectivity of random fault networks following a power law fault length distribution. Water Resour Res 33(7):1567–1583

Bour O, Davy P (1998) On the connectivity of three-dimensional fault networks. Water Resour Res 34(10):2611–2622

Bowden A, Rigg A (2004) Assessing risk in CO_2 storage projects. Aust Petrol Prod Explor Assoc J 44(1):677–702

Carbotte SM, Macdonald KC (1994) Comparison of sea-floor tectonic fabric at intermediate, fast, and super fast spreading ridges—influence of spreading rate, plate motions, and ridge segmentation on fault patterns. J Geophys Res Solid Earth 99(B7):13609–13631

Cowie PA, Scholz CH, Edwards M, Malinverno A (1993) Fault strain and seismic coupling on Midocean ridges. J Geophys Res Solid Earth 98(B10):17911–17920

Cowie PA, Sornette D, Vanneste C (1995) Multifractal scaling properties of a growing fault population. Geophys J Int 122(2):457–469

Davy P (1993) On the frequency-length distribution of the San-Andreas fault system. J Geophys Res Solid Earth 98(B7):12141–12151

Dershowitz WS, Einstein HH (1988) Characterizing rock joint geometry with joint system models. Rock Mech Rock Eng 21(1):21–51

Dou C, Woldt W, Bogardi I (1999) Fuzzy rule-based approach to describe solute transport in the unsaturated zone. J Hydrol 220(1–2):74–85

Espie T (2004) Understanding risk for the long-term storage of CO_2 in geologic formations. In: Seventh international conference on greenhouse gas control technologies, Vancouver, Canada

Gudmundsson A (1987) Geometry, formation and development of tectonic fractures on the Reykjanes Peninsula, Southwest Iceland. Tectonophysics 139(3–4):295–308

Gueguen Y, Dienes J (1989) Transport-properties of rocks from statistics and percolation. Math Geol 21(1):1–13

Hundecha Y, Bardossy A, Theisen HW (2001) Development of a fuzzy logic-based rainfall-runoff model. Hydrol Sci J 46(3):363–376

IPCC (Intergovernmental Panel on Climage Change) (2005) In: Metz B, Davidson O, de Coninck H, Loos M, Meyer L (eds) Special report on CO_2 capture and storage. Cambridge University Press, UK, pp 208–210

Kagan YY (1997) Seismic moment-frequency relation for shallow earthquakes: regional comparison. J Geophys Res Solid Earth 102(B2):2835–2852

Kumar V, Schuhmacher M, Garcia M (2006) Integrated fuzzy approach for system modeling and risk assessment. In: Modeling decisions for artificial intelligence, pp 227–238

Laherrere J, Sornette D (1998) Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. Eur Phys J B 2(4):525–539

Liu HH, Bodvarsson GS, Finsterle S (2002) A note on unsaturated flow in two-dimensional fracture networks. Water Resour Res 38(9)

Main I (1996) Statistical physics, seismogenesis, and seismic hazard. Rev Geophys 34(4):433–462

Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with fuzzy logic controller. Int J Man Mach Stud 7(1):1–13

Masihi M, King PR, Nurafta P (2006) Effect of anisotropy on finite-size scaling in percolation theory. Phys Rev E 74(4)

Masihi M, King PR, Nurafta P (2007) Fast estimation of connectivity in fractured reservoirs using percolation theory. SPE J 12(2):167–178

Mays MD, Bogardi I, Bardossy A (1997) Fuzzy logic and risk-based soil interpretations. Geoderma 77(2–4):299–315

McBratney AB, Degruijter JJ (1992) A continuum approach to soil classification by modified fuzzy K-means with extragrades. J Soil Sci 43(1):159–175

McBratney AB, Odeh IOA (1997) Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77(2–4):85–113

Novak V, Perfilieva I (2000) Discovering the world with fuzzy logic. Physica-Verlag/Springer, Heidelberg/New York

Odeh IOA, McBratney AB, Chittleborough DJ (1992) Soil pattern-recognition with fuzzy-C-means—application to classification and soil-landform interrelationships. Soil Sci Soc Am J 56(2):505–516

Oldenburg CM, Bryant SL, Nicot JP (2008) Certification framework based on effective trapping for geologic carbon sequestration. Int J Greenhouse Gas Control (in review)

Panigrahi DP, Mujumdar PP (2000) Reservoir operation modelling with fuzzy logic. Water Resour Manage 14(2):89–109

Priest SD, Hudson JA (1981) Estimation of discontinuity spacing and trace length using scanline surveys. Int J Rock Mech Min Sci 18(3):183–197

Pruess K (2008) On CO_2 fluid flow and heat transfer behavior in the subsurface, following leakage from a geologic storage reservoir. Environ Geol 54(8):1677–1686

Renshaw CE (1999) Connectivity of joint networks with power law length distributions. Water Resour Res 35(9):2661–2670

Robinson PC (1983) Connectivity of fracture systems—a percolation theory approach. J Phys A Math Gen 16(3):605–614

Rouleau A, Gale JE (1985) Statistical characterization of the fracture system in the Stripa Granite, Sweden. Int J Rock Mech Min Sci 22(6):353–367

Savage D, Maul P, Benbow S, Walke R (2004) A generic FEP database for the assessment of long-term performance and safety of the geological storage of CO_2. http://www.co2captureandstorage.info/docs/QuintessaReportIEA.pdf

Scholz CH, Cowie PA (1990) Determination of total strain from faulting using slip measurements. Nature 346(6287):837–839

Segall P, Pollard DD (1983) Joint formation in granitic rock of the Sierra-Nevada. Geol Soc Am Bull 94(5):563–575

Shrestha BP, Duckstein L, Stakhiv EZ (1996) Fuzzy rule-based modeling of reservoir operation. J Water Resour Plan Manage Assoc 122(4):262–269

Sornette D, Sornette A (1999) General theory of the modified Gutenberg–Richter law for large seismic moments. Bull Seismol Soc Am 89(4):1121–1130

Stauffer D, Aharony A (1992) Introduction to percolation theory. Taylor & Francis, London

Takagi T, Sugeno H (1985) Fuzzy identification of systems and its application for modeling and control. IEEE Trans Syst Man Cybern 15(1):116–132

Uricchio VF, Giordano R, Lopez N (2004) A fuzzy knowledge-based decision support system for groundwater pollution risk evaluation. J Environ Manage 73(3):189–197

Wildenborg T, Leijnse A, Kreft E, Nepveu M, Obdam A, Orlic B (2005) Risk assessment methodology for CO_2 storage: the scenario approach. In: Carbon dioxide capture for storage in deep geologic formations. Elsevier, Amsterdam

Wildenborg T, Leijnse T, Kreft E, Nepveu M, Obdam A (2004) Long-term safety assessment of CO_2 storage: the scenario approach. In: Seventh international conference on greenhouse gas control technologies, Vancouver, Canada

Zhang K, Wu Y, Bodvarsson GS, Liu HH (2004) Flow focusing in unsaturated fracture networks: a numerical investigation. Vadose Zone J 3:624–633

Zweigel P, Lindeberg E, Moen A, Wessel-Berg D (2004) Towards a methodology for top seal efficacy assessment for underground CO_2 storage. In: Seventh international conference on greenhouse gas control technologies, Vancouver, Canada