J. Ind. Geophys. Union (January 2016)
Special Volume-1/ 2016 pp: 30-36

Visualizing Uncertainty in CO_2 Plume Migration During Sequestration

S. Srinivasan* and H. Jeong

Department of Petroleum and Geosystems Engineering, University of Texas at Austin, 200 E Dean Keaton, C0300, Austin, TX 78712

*Corresponding Author: sanjay.srinivasan@engr.utexas.edu

ABSTRACT

During the operation of a geological carbon storage project, verifying that the CO_2 plume remains within the permitted zone will be of particular interest both to regulators and operators. A model selection algorithm was developed, which refines an initial suite of subsurface models representing the prior uncertainty to create a posterior set of subsurface models that reflect injection performance consistent with that observed. Such posterior models can be used to represent uncertainty in the future migration of the CO_2 plume. The method provides a very inexpensive alternative to map the migration of the plume and the associated uncertainty in migration paths due to the fact that only injection data is required. An essential aspect of the model selection algorithm is to group prior models on the basis of their connectivity. The base algorithm assesses that connectivity using a physical proxy such as a random walker. An alternate approach would be to use statistical tools for assessing connectivity of models. In this paper, we present an approach to compute the shortest connected path between well locations in an aquifer model and to define a measure of similarity of shape based on the concept of a discrete Fréchet distance.

INTRODUCTION

A key difficulty in accurately predicting the CO_2 plume migration path in the subsurface is the uncertainty in the underlying geology. To resolve that uncertainty, the observed flow response of the subsurface system can be used to optimally constrain the subsurface models. The key underlying technology is thus parameter estimation (commonly referred to as history matching). As opposed to traditional permeability or probability perturbation methods (e.g. Pro-HMS - (Kim, 2008)) that work on updating permeability values at a pixel-level, the proposal is to work directly in the realization space of the random function (RF) model. In traditional approaches, P (A|B) assumes that A is a simulation event at a pixel level (e.g. permeability at a location u) and B is the available data. In this new proposal, A will be a model realization of the RF (i.e, it will be all the pixels taken jointly). Thus P(A|B) will be the conditional probability of a model given the data. Similarly, P(A|C) is the probability of a model given the injection data C. At the end of the merging process, we get P(A|B,C) i.e. the probability that a model A that exhibits conformance to the available static and injection data. If the distribution P(A|B,C) exhibits a peaked structure, this implies that the available data cause some of the models to be preferred over others. The history matching (injection data integration) approach therefore becomes an exercise in selecting the most plausible model(s) guided by the available data.

Here we propose a new perspective: rather than estimating the properties of the storage formation, we seek to infer the location of the plume of injected CO_2 within that formation. In particular, we wish to be able to identify deviations of the plume from its anticipated migration path, as these deviations can have substantial regulatory ramifications. A primary cause of such deviations will be heterogeneities in the storage formation (baffles, sealing faults, high permeability streaks, and the like), which are not known at the beginning of the injection. A unique model-selection algorithm will be implemented that uses the injection data to select a subset of initial aquifer models that best adhere to observed injection characteristics. The simulation of plume displacement in this reduced subset of models yields the uncertainty in the plume location.

MODEL SELECTION

In this new framework, we start with a whole suite of plausible models for the particular subsurface system being studied. This set of prior models could be rather large, reflecting uncertainty in depositional environments and aquifer architecture. This is likely to be the situation at the outset of many geologic sequestration projects. This suite of prior models will be processed through a fast transfer function model such as a streamline simulator or a random walker or alternatively, a fast algorithm for assessing reservoir connectivity as in this paper. The fast transfer function model is a proxy for the aquifer flow model and is only used to gauge rapidly the comparative performance of the prior models. Denoting the response of the proxy function on the prior model $z^l(u), \forall u \in$ Domain as

$f^l(\mathrm{u})$ and assuming that there are L prior models, we can calculate a distance matrix:

$$D = \begin{bmatrix} d_{11} & \cdots & d_{1L} \\ \vdots & \ddots & \vdots \\ d_{L1} & \cdots & d_{LL} \end{bmatrix}$$

the distances d_{ij} are defined as $d_{ij} = ||f^i(\mathrm{u}) - f^j(\mathrm{u})||$. Multivariate classification techniques such as Principal Component Analysis, Independent Component Analysis or k-mean cluster analysis would yield a grouping of aquifer models in terms of similarity of the proxy function response. We emphasize that this grouping procedure might result in geological models representing different environments/architecture getting grouped together. This would indicate that despite the overt differences in geology, those models exhibit similar connectivity characteristics that have a predominant influence on the observed injection data. This feature will also be exploited below to provide a quantitative estimate of uncertainty in the plume location.

Based on the distances between the models that make up a cluster, an average realization representative of the cluster can be computed by performing distance-weighted averaging of the realizations making up the cluster.

Alternatively, any of the models in a cluster could be retrieved as a representative model for that cluster. We denote these representative models as $z^m(\mathrm{u}), m=1,...M$, where M is the number of clusters identified. Since each of the M models is equally conditioned to the available well information (such as log/core data etc.), each one is equally plausible. In that case, the prior probability $p(z^m(\mathrm{u})) = 1/M$.

Flow simulations with the full physics of the flow process of interest can be performed on the M reduced models. Denoting the corresponding responses as $RF^m(\mathrm{u}), m=1,..,M$ and given the observed injection history RF_{ref}, we can calculate the deviation $\sigma_m^2 = ||RF_{ref} - RF^m(\mathrm{u})||$. If $RF^m(\mathrm{u})$ is a vector in time i.e. $\mathrm{RF}^m(\mathrm{u})$, then the deviation is calculated as $\sigma_m^2 = [\mathrm{RF}_{ref} - \mathrm{RF}^m(\mathrm{u})]^T [\mathrm{RF}_{ref} - \mathrm{RF}^m(\mathrm{u})]$. Assuming a Gaussian probability distribution for the mismatch with the mean as RF_{ref} and the variance as σ_m^2, we can calculate the probability that any model in the cluster m would match the observed data i.e. $p(RF_{ref}|z^m(\mathrm{u}))$. The prior probability of the response RF_{ref} within the pool of M responses for the reduced models i.e. $p(RF_{ref})$ can also be calculated by pooling together the responses for the M reduced models and finding the probability corresponding to RF_{ref}.

The objective is to derive the posterior probability $p(z^m(\mathrm{u})|RF_{ref})$ i.e. the updated probability for the M models given the observed injection response. This can be computed by a straight forward application of Bayes' rule:

$$p(z^m(\mathbf{u})|RF_{ref}) = \frac{p(RF_{ref}|z^m(\mathbf{u}))}{p(RF_{ref})} \times p(z^m(\mathbf{u}))$$

All the quantities on the RHS are available. Hence, the updated probability can be calculated. A random draw can be made from the updated probability to retrieve the particular model $z_i^m(\mathrm{u})$, where the subscript i denotes the iteration for the application of this workflow.

After an application of the above process, a group m that exhibits the flow response that is similar to the observed injection data is identified. It is quite likely that a one-time application of this process would not yield a satisfactory match to the observed injection history. Consequently, the process is repeated using the member models that make up the group m. Multivariate classification (cluster analysis or PCA) is performed in order to further subdivide the member of group m. Application of the Bayes' rule now will yield the subgroup $z_i^{m'}(\mathrm{u})$ for iteration $i=2$. Repeated application of the process therefore refines further the selection of the model that is closest to the observed injection history. The entire process is summarized in the schematic in Figure 1. Flow chart of the model selection process using injection data..

A DEMONSTRATION EXAMPLE

This approach is applied to a synthetic case, which is modeled after the In Salah field. The preferential orientation of the high permeability features in the In Salah field is in the SE-NW direction - so we assume that the major orientation of the channel reservoir is in that direction. Fig. 2 shows the preferential orientation, a training image and two realizations generated using the multiple point stochastic simulation algorithm SNESIM (Strebelle, 2002). 400 realizations are generated using SNESIM and then they are clustered based on the similarity in connectivity characteristics of the models.

We need a direct measure of similarity of spatial connectivity in order to cluster reservoir models by spatial connectivity. We introduce the notion of a connected path to describe spatial connectivity of reservoir models. Since the injected fluid flows along permeable zone, the connected path is defined as the most permeable path between an injection point and a target point (which could be a hypothetical monitoring location) so as to delineate reservoir connectivity. Details of the procedure to compute the connectivity can be found in Srinivasan and Jeong (2012). Broadly, the computation procedure consists of identifying the shortest connected path between an injector and a hypothetical producer and then computing the distance between the shortest path for two different reservoir models using the notion of the Discrete Fréchet Distance (DFD – Eiter and Mannila, 1994). This permits the computation of the distance matrix. The realizations are projected onto a metric space or a kernel space using multidimensional scaling (MDS – Romney, 1972; Kruskal, 1978) or kernel principal component analysis (KPCA -

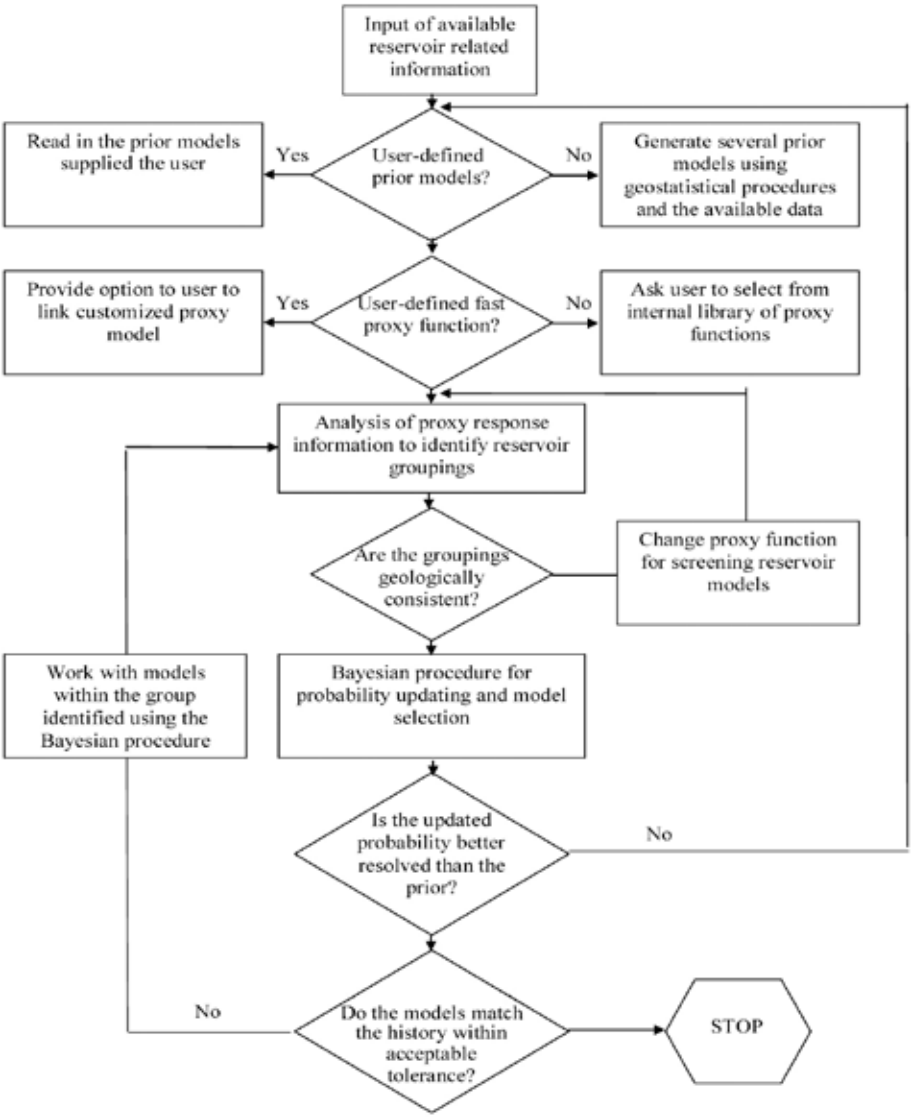


Figure 1. Flow chart of the model selection process using injection data.

Scholkopf, 1999) of the distance matrix. In the projected space, the realizations are clustered using K-means clustering.

In this synthetic example, there is one producer and three injectors. The region of interest is in the vicinity of the three injectors: KB-501, KB-502, and KB-503. We calculate the connected paths between the producer and the three injectors over 400 realizations as shown in Figure 3. Connected paths between a producer and three injectors.. The distance between the models is calculated in terms of the Fréchet distances. The distance matrix of five realizations for the path from KB-501 to PROD is given in Table 1. The discrete Fréchet distances between five models for the path from KB-501 to PROD.. Models #1 and #5 look similar and #2, #3, and #4 also look similar. However, the two groups have quite different shapes. Thus, the DFDs between the same group members are small, but the DFDs between the different group members are large. After we calculate all of the DFDs for KB-501 to PROD, KB-502 to PROD, and KB-503 to PROD, the combined distances are calculated as shown in Figure 4. An example for computation of the combined distance.. Then the realizations are clustered according to the combined distances between the realizations.

Multi-Dimensional Scaling (MDS) requires high dimension as well as high computational cost for transforming the distances to the metric space. For these reasons, MDS is not appropriate for projection of the 400 models. However, MSD is beneficial for the visualization of the relation between models. Therefore we projected the 20 medoids into 2D space using MDS so that their relation is visualized as shown in Figure 5. Two-dimensional

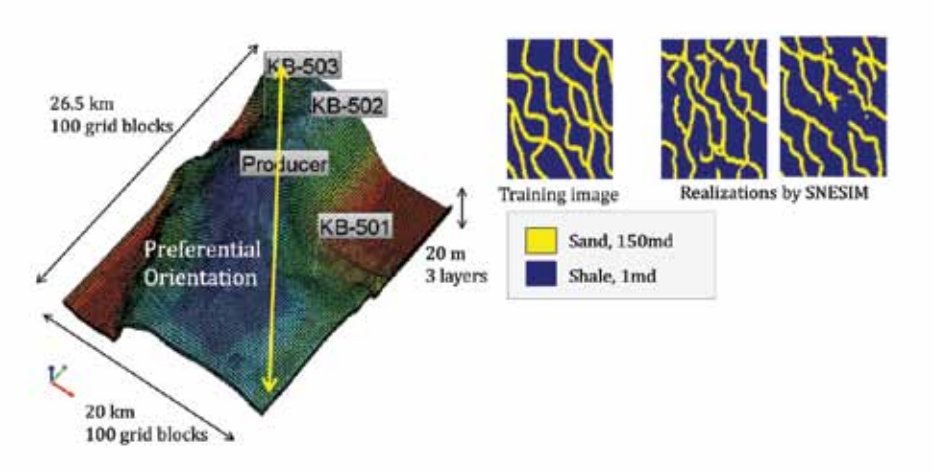


Figure 2. Preferential orientation of reservoir connectivity and a training image of the In Salah field. The color scale shown in the Fig. corresponds to the reservoir top.

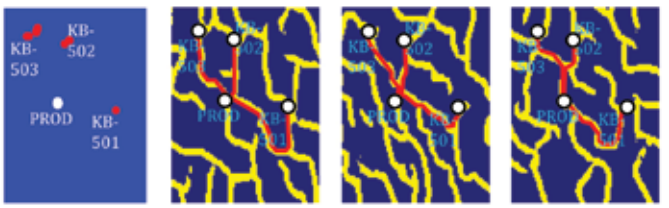
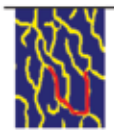
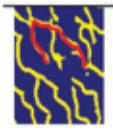
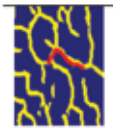
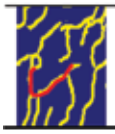


Figure 3. Connected paths between a producer and three injectors.

Table 1. The discrete Fréchet distances between five models for the path from KB-501 to PROD.

Discrete Fréchet Distance		#1	#2	#3	#4	#5
	#1	0.00	39.00	37.34	49.93	15.81
	#2	39.00	0.00	33.94	25.18	39.00
	#3	37.34	33.94	0.00	34.21	23.02
	#4	49.93	25.18	34.21	0.00	49.93
	#5	15.81	39.00	23.02	49.93	0.00

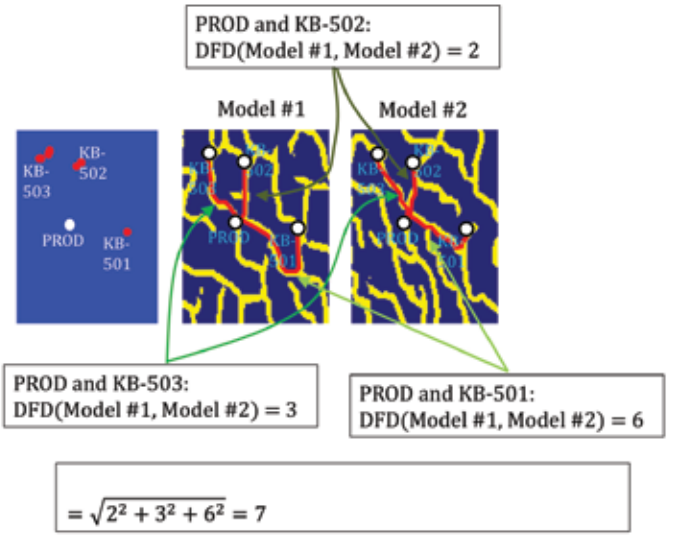


Figure 4. An example for computation of the combined distance.

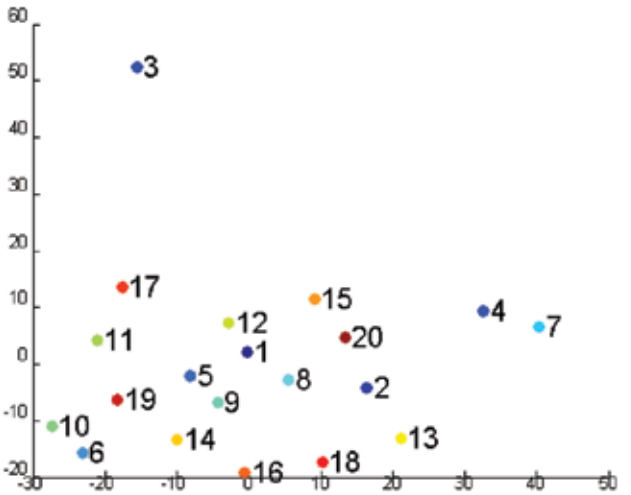


Figure 5. Two-dimensional space of 20 medoids projected by MDS.

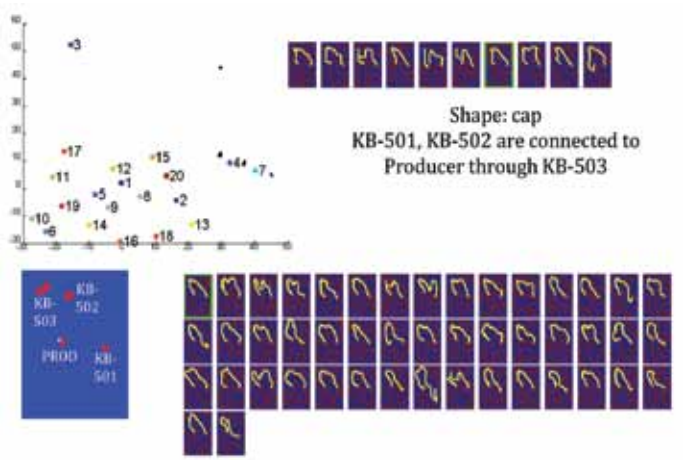


Figure 6. Comparison of cluster #4 and cluster #7.

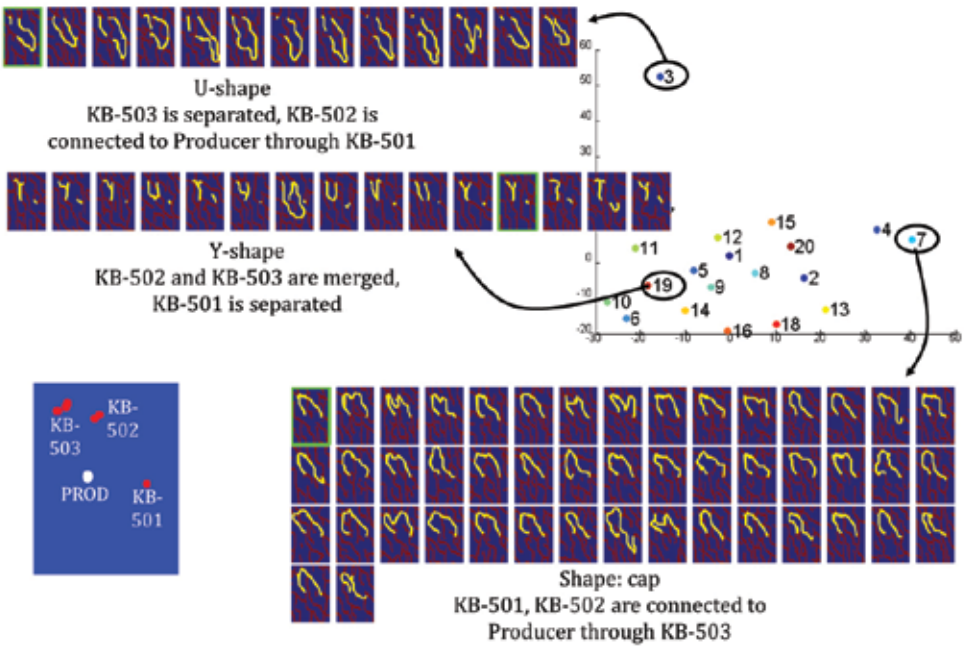


Figure 7. Comparison of cluster #3, cluster #7, and cluster #19

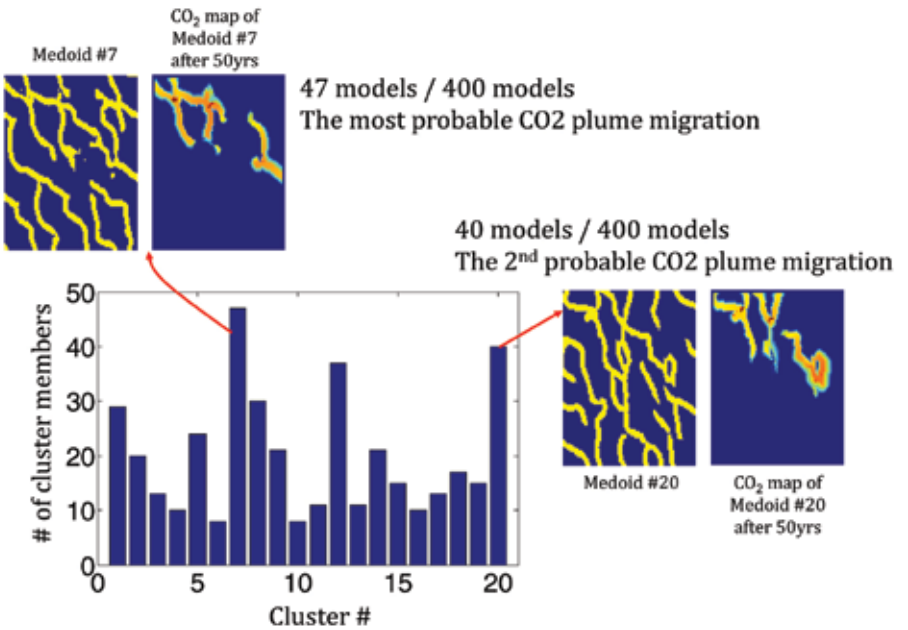


Figure 8. The extracted features of CO_2 plume migration.

space of 20 medoids projected by MDS.. This is only for visualization of the 20 medoids' locations and MDS is not used to cluster the 400 models.

Figure 6. Comparison of cluster #4 and cluster #7. shows the connected paths and the facies map of cluster #4 and cluster #7. The yellow lines in Figure 6. Comparison of cluster #4 and cluster #7. are the connected paths. Since cluster #4 and cluster #7 are very close to each other, shape of their connected paths should be similar. The shape of their connected paths looks like a cap. KB-501 and KB-502 are connected to the producer through KB-503.

Figure 7. Comparison of cluster #3, cluster #7, and cluster #19 shows the connected paths and the facies map of cluster #3, #7, and #19. They have long distances to each other, so their shape of the connected paths is quite different.

Since we assume the connected paths are similar to CO_2 plume migration paths, we do not need to run a simulator to quantify uncertainty of the CO_2 plume migration. We cluster the models based on the shape of the connected paths instead of the simulation results of CO_2 plume migration to quantify the uncertainty of CO_2 plume migration. Once we cluster the models, we are able to extract the features CO_2 plume migration by simulating the representative models instead of all the models. Figure 8. The extracted features of CO_2 plume migration. shows the two extracted features of CO_2 plume migration. In Figure 8. The extracted features of CO_2 plume migration., the histogram shows the number of cluster members for each cluster. Since cluster #7 has the most number of cluster members, the CO_2 map of cluster #7 can be considered as the most probable CO_2 plume migration. Likewise, the CO_2 map of cluster #20 can be considered as the second probable CO_2 plume migration.

CONCLUSIONS

The paper presents a novel approach to classify prior geologic models on the basis of connectivity of flow paths within the reservoir. Dissimilarity between prior models is computed on the basis of the differences in the characteristics of the shortest connected path between well pairs. The concept of discrete Frechet distance is used for this computation. The dissimilarity measures are grouped using k-medoid clustering. Flow simulation on models belonging to any cluster reveals similarity in CO_2 swept regions in all models belonging to a cluster. However, the current approach has two problems in quantifying the uncertainty of CO_2 plume migration. The first problem is that CO_2 may migrate along disconnected paths, in other words local flow and transport gradients may cause the CO_2 plume to bridge permeability discontinuities. The current approach compares only connected paths. We therefore should measure path similarity regardless of connection between points of interest. The second problem is that the current approach can be applied to only convection-dominant models such as channel, fractured reservoir models. Thus, development of a similarity measure applicable to general geologic models is necessary.

REFERRENCES

Eiter, T., and Mannila, H., 1994. Computing discrete Fréchet distance. Technische Universitat Wien Technical Report CD-TR, v. 94, pp: 64.

Srinivasan, S. and Jeong, H., 2012. Modeling the Uncertainty in CO_2 Plume Migration During Sequestration Using a Model Selection Approach. Proceedings of the Ninth Conference on Geostatistics for Environmental Applications, geoENV2012, Valencia, Spain, September 19–21.

Kim, Y.H., Srinivasan, S., and Bryant, S.L., 2008. Efficient Region Decomposition for Parallel Simulation Accounting for Uncertainty in the Reservoir Model. SPE 116752, SPE Annual Technical Conference and Exhibition, Denver.

Kruskal, J. B., and Wish, M., 1978. Multidimensional Scaling. Sage University Paper series on Quantitative Applications in the Social Sciences, Sage Publications, Newbury Park, CA, pp: 07-011.

Romney, A. K., Shepard, R. N., and Nerlove, S. B., 1972. Multidimensional Scaling - Theory and Applications in the Behavioral Sciences. Seminar Press., New York v.2.

Scholkopf, B., Smola, A., and Muller, K.R., 1999. Kernel Principal Component Analysis. In B. Scholkopf, C. J. C. Burges, and A. J. Smola, editors, Advances in Kernel Methods–Support Vector Learning, MIT Press, Cambridge, MA, pp: 327-352.

Strebelle, S., 2002. Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics, Mathematical Geology, v.34, 1.