Paper 126, CCG Annual Report 14, 2012 (© 2012)

Clustering as an Alternative to Ranking Realizations

Saina Lajevardi, and Clayton V. Deutsch

Although a large number of realizations (typically 100 or more) are required to understand the uncertainty at a geological site, processing all of the realizations through a complex transfer function is not practical. Despite the fact that realizations are sufficiently different to allow us to understand the uncertainty, they still carry large spatial correlation due to common conditioning data and modeling parameters. This enables practitioners to consider fewer realizations. Ranking has been around for a while as a widely practiced approach to select a few realizations for detailed processing. In this paper, we tackle this problem through a different paradigm, that is, clustering the realizations according to relevant geological and spatial characteristics. There is no a-priori assumption about whether the realizations will perform better or worse. The result of clustering is to partition the realizations into different groups that share similar features. The results could be expressed by a few representative (e.g. centroid) values associated to every cluster. This approach is different, but complementary to the conventional ranking approach.

Introduction

Stochastic simulation has been around as one of the earliest techniques in risk and sensitivity analysis in many fields as well as spatial sampling in earth sciences. Generating a large number of realizations by Monte Carlo Simulation (MCS) can be thought of as simple random sampling. This is the most primitive approach in stochastic simulation because of the large number of realizations that must be processed to understand response uncertainty. Selective sampling was developed to require fewer of realizations. Algorithms such as Latin Hypercube Sampling (LHS) or orthogonal sampling could also be used for this purpose. The challenge with this technique in geostatistics is that the geological models have very high dimension - often tens of millions of grid cell locations and selecting a quantile a-priori is not possible.

The solution adopted in geostatistics has been to rank the realizations by a quick-to-calculate measure also known as a simple transfer function. It is essential that the simple transfer function be highly correlated to the full transfer function that is of actual interest. A large number of realizations are generated, then a few realizations are chosen for more detailed analysis based on their quantile position using the simple transfer function. The ranking changes with any change in the analysis, such as when the area of interest changes, the well locations change, the recovery process changes and so on.

This paper presents a different approach; the starting point is the same: a large number of realizations, then the realizations are grouped into clusters. Clustering has been around as a classical procedure for data description in data mining and data analysis. The concept of clustering is in many ways close to the data selection as it attempts to partition data into different groups. Every group could then be described by a representative (e.g. centroid). In this context, the representatives of the clusters could be considered as the selected realizations which would undergo the further processing in reservoir assessment.

Clustering the realizations could be applied in different concepts. Every realization could be considered as a long array of grid cells describing a geological feature of the reservoir. This is a very close problem to multi-class clustering. In that case, every realization is linearized to be a vector. For example, if every realization contains $100 \times 10 \times 10$ grid cells, every vector in the multi-class matrix will have a length of 10^4. If clustering is to be applied on 100 realizations, the matrix has a dimension of 100×10^4. Kernel can then be applied to the realizations and measure the similarities between the data points (grid nodes). This way, the grouping is applied over the grids than realizations. In our example, the kernel matrix will be of dimension 100×100 (including the dot products of every incident).

Here, however, our approach to clustering is slightly different. We consider every realization as the collection of data points and measure the similarities between the realization through some features associated to every realization. Because of the uncertainty exists in the estimation of the realizations, every realization describes the geological (or flow) features slightly different than others. However, depending on the reservoir distribution at every realization, feature's measurements are different for different realizations. Depending on how similar the distribution of reservoir is in the realizations, features might reveal higher (lower) correlations. The measured similarities between features of realizations is the criteria to group them together. Realizations in one group are expected to perform closely in the further processes; this could also be considered as a way

to reduce the number of realizations required for further processes.

Clustering: Different Paradigm to Selecting Realizations

Clustering implies grouping number of objects in a way that the objects in one group are more similar to each other compared to the ones in other groups. There are many algorithms for data clustering in the literature. Since evaluating the clustering performance is not our purpose at this paper, we focus on a reasonable, simple clustering algorithm to group the realizations. The k-means (centroid-based) clustering is a simple form of clustering that is widely used. K-means clustering is based on measuring the distance between the realizations considering all dimensions (features). This is very close to vector quantization, where the centroid are randomly chosen from a codebook (the codebook in our case is the collection of realizations). This is basically an iterative clustering approach. The algorithm however, lacks the power to extract the natural cluster of the data; it is required to decide on the number of groups beforehand. That is not the only disadvantage that K-means clustering suffers from. K-means clustering applies partitioning on data linearly. This in fact limits the performance of clustering and is not the most appropriate approach most of the time; K-means clustering does not perform well on data having nonlinear structure.

In the K-means clustering, the number of clusters K should be determined prior to data clustering. This typically is considered as one of the disadvantages of K-means approach. There are several work in literature concentrating on this issue and suggesting global K-means clustering to resolve the initialization problem incrementally (Tzortzis and Likas, 2008). However, we avoid the complexity at this point and devote this work to the main concept of Kernel K-means clustering over the realizations. We also determine the numbers of clusters based on the numbers of principal components of the data set. Principal component analysis in multivariate statistics is widely known as an effective dimension reduction method which represents data at its most variate dimensions. Since sufficiently large principal components represent effective dimensions of the data set, they might be a good approximation of the number of clusters. In addition, we should consider the fact that large number of clusters are not required at this work. As the number of selected realizations are going to be process further, smaller numbers of them which can efficiently represent the realizations set are more appreciated.

Kernel K-means Clustering

In this work, we intend to cluster the realizations according to several features in the multi-dimensional space. As was discussed, this is to allow other understanding of realizations in addition to the common techniques in practice such as ranking. The main purpose of ranking is to reduce the number of realizations for further processes. We propose realizations clustering for the same purpose; every cluster is the collection of realizations which are more similar and could be represented with one of them (e.g. centroid). The selected realizations go through further processes instead of using all realizations which is almost always impractical.

In ranking, the realizations are sorted based on the measure of one feature while in our clustering, grouping data based on one feature does not make sense or more precisely is just equivalent to ranking. Therefore, we apply clustering in a multi-dimension domain. Every dimension is one feature which has been measured for all the realizations. Note that the features are not always structured linearly, this necessities the utilization of a clustering algorithm which is not confined to partitioning data only linearly. In other words, the clustering algorithm which partition data considering the nonlinearity structure is the most adequate approach.

Kernel function ensures that the measured similarities between the realizations captures the effect of nonlinear features as well. When kernel is applied to the data matrix, the data is mapped from input space to the feature space. This transformation is nonlinear and has larger dimensions than data space. In the feature space, since the nonlinear structure of data has been capture, applying a clustering algorithm which partition data linearly no longer would be a problem. This brings us back to the utilization of the most common, conventional clustering algorithm, K-means clustering. K-means clustering is then separate data into non-overlapping clusters. It groups data based on their closeness to the center of the group. The criteria for closeness is Euclidean distance which is applied in multi-dimensional space in our scenario (more than two features are suggested to be included).

The given data set $\chi = \mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_N}$ is to be partitioned into K clusters of $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_N$ and every $x_n \in \mathbf{R}^d$. In K-means clustering, when a vector belongs to a cluster, its Euclidean distance to the center of that cluster should be the smallest with respect to other centroid. In other words, the data (vectors) with more

similarities should be grouped together. The clustering error for the entire data set is given by

$$E(m_1, \cdots, m_K) = \sum_{n=1}^{N} \sum_{k=1}^{K} I(x_n \in \mathcal{C}_k) \|x_n - m_k\|^2 \tag{1}$$

where I indicates the existence of a vector in a specific cluster. We apply K-means clustering in the feature space $\mathcal{F}(n \times n)$, where the kernel function $K_{ij} = \phi(x_i)^T \phi(x_j)$ has transferred the data from original space $\mathbf{R}^{n \times p}$ into higher dimension of feature space. The simplest transfer function ϕ we apply here is $K(x_i, x_j)$ which directly provide the inner products in feature space. The first step in kernel K-means clustering is to stabilize the centroid of the cluster. The initial centroid are chosen randomly. This requires the algorithm to look for the minimum distance between every kernel data $\phi(X_n)$ and all centroid of clusters $\|\phi(X_n) - m_i\|^2$ where $1 \leq i \leq k$. The cluster which is equivalent to $\arg\min_i(\|\phi(X_n) - m_i\|^2)$ includes $\phi(X_n)$. Next to the addition of every kernel data to a cluster, the centroid of the cluster is updated by averaging the belonging kernel data of the corresponding cluster. This continues until the centroid are no longer changing and every realization belongs to its group. At the end of this procedure, every centroid is equivalent to $m_k = \frac{\sum_{n=1}^{N} I(x_n \in \mathcal{C}_k)\phi(x_n)}{\sum_{n=1}^{N} I(x_n \in \mathcal{C}_k)}$. The objective of this approach is to minimize the clustering error in feature space where the data vector is replaced by its kernel transform.

$$E(m_1, \cdots, m_K) = \sum_{n=1}^{N} \sum_{k=1}^{K} I(x_n \in \mathcal{C}_k) \|\phi(X_n) - m_i\|^2 \tag{2}$$

This algorithm is believed to converge when kernel matrix is positive semidefinite (PSD). The kernel matrix which directly apply the inner products of data set is always positive semidefinite and suitable at this stage. There are different characterizations associated with PSD matrices. One is that all eigenvalues are requires to be real and positive (always true with the inner product matrix). Since in our case, the kernel matrix is generated by inner product of all data values, its positive definiteness is guaranteed.

The purpose of clustering the realizations is to identify a smaller number of realizations that would perform differently. Kernel clustering is a reasonable choice due to the fact that variables measured as realizations properties might inherit nonlinear structure which will not be caught if conventional clustering techniques are applied to the data space. For example, connected hydrocarbon volume (CHV) is a feature which its determination not only depends on the net distribution of realization but well placement as well. If it is considered in the clustering (which it should due to its importance in recovery), its nonlinear structure would not be captured in the K-means realizations and the clustering results are not accurate, since it treats CHV as if it is linear. Kernel exploits data nonlinearity structure and reveals it in the feature space. Therefore, applying the linear clustering i.e. K-means on the feature space enables the realizations with more similarities group together regardless of the nonlinear structure.

Our experiment shows that CHV has a nonlinear structure compared to the other features such as net volume, numbers of geoobjects and tortuosity. Thus, it dominates the clustering and therefore, the realizations in different group show very small overlap in terms of CHV. This could be another way to look at ranking of CHV.

The choice of features depends on the data available and an understanding of what might be important for a particular problem. For example, net volume can mostly be considered as a linear variable specially when the heterogeneity of the realizations are not high. The purpose of applying kernel based clustering is to detect the nonlinear structure of data and enable data clustering in a higher dimension (kernel dimension).

Experiments

Consider 100 realizations modeling a small area of 100 grids in X direction, 20 in Y direction and 100 in Z direction. Every realization is associated with several attributes that we can compute quickly. Variables such as (1) the net volume (2) the numbers of geoobjects (3) the effective permeability (4) the tortuosity of the largest object (5) local connectivity; connected hydrocarbon volume considering well placement. Some of these features are closely related and some less. The correlation between the features also vary depending on the heterogeneity of the realizations.

Some other measurements could directly result from the technology that is applied for recovery (e.g.

SAGD). Local connectivity is another feature that appears to be different depending on where the recovery well pair is planned to be placed. As connected and uniform the reservoir is simulated to be in a realization, The number of geoobjects is smaller and the volume of each is larger. Even at this stage, the Image cleaning techniques such as erosion/ dilation could significantly influence the number of geoobjects depending on how intense the erosion/ dilation are applied. Local connectivity is also to be considered as another realization's feature. How connected the net area is around the well placement would result in different categorization of the realizations. Typically, the reservoir distributions are not uniform and the local connectivity is crucial to well placement and recovery performance.

The numbers of connected well pairs throughout the reservoir could be considered as another important feature of realization. Having set the well placement, different realizations appear to be connected differently from well to well depending on the net distribution of the reservoir. Evaluating the numbers of cells connecting specific number of wells is an effective factor in recovery.

Spatial features of the realizations such as the entropy, specific property of the variogram (the distance where the variogram (vertical) reaches 50% of its sill).

Some more measurements associated with the flow could be evaluated and used to help distinguish realizations accordingly. Fast flow simulation or proxy models could describe the realizations differently. Similar to the other measurements related to the geology of the realizations, these properties could also be seen as part of the clustering features. The example in this paper only employs a facies model and provide measurement factors such as net volume, geometry (number of geoobjects), tortuosity of the biggest objects, specific distance of variogram and connected hydrocarbon volume for the entire reservoir where the well placed at the center. A small example illustrates the approach.

Sensitivity Analysis

Applying the clustering algorithm, it is important to study how stable the resulting clusters are if noise is added to the data. Note, that this would be different based on the available measurements and how correlated they are. Also, the nonlinearity structure of the data in the feature domain could have a large influence on the way the clustering works, the interpretations, the stability and the results obtained.

In the first example, the realizations have large net volume distribution. We have added noise as one of the variable to our data set and applied both the clustering algorithm to decide on how the clusters change. Note that we keep the number of clusters the same before and after the addition of noise so that the difference in clustering is easier to understand. We have applied the kernel clustering to the data set before and after the addition of noise variable which result in the average error of E $= 4.09310^{-7}$ at both situations. Also, the average distance between the centers is E $= 1.4610^{-6}$ which is bigger than the distance inside the cluster (expected). The number of clusters could be adjusted if this result is not reasonable. Every realization is also grouped within the same cluster with and without the noise; the similarity of clusters realizations is 100%. However, applying K-means in data space when there is no noise added, shows clustering error of 0.1161, while it changes to 0.2469 for the case when noise is added. As can be seen, the clustering error is much larger for the case when K-means clustering is applied to the data space instead of feature space. This confirms that the nonlinearity structure in the data space cannot be exploited using K-means clustering. Also, the similarity between clusters before and after the addition of noise is 25% which is the sign of instability. This is as opposed to Kernel K-means clustering in our experiment which shows sign of stability against noise. Figures 1, 2, and 3 demonstrates the result of kernel K-means clustering on data in the original space. CHV seems to dominate the clustering in this example. The scatter plots shows more structured with the features such as net volume, small area CHV, or effective permeability and less correlation (effect of) tortuosity or number of geoobjects. This is why kernel clustering looks proper at this point; data in feature space reveals the strength of the effect of every variable in clustering.

Our analysis shows that the variable which has the least correlation with the noise most times (87% in our case), while it has sufficiently large correlation with other variables and hence ir dominate the clustering process. For example, the following table demonstrates the correlation coefficients of three feature vectors for 100 realizations. The second table 2 demonstrates the correlation coefficients when another variable of tortuosity has been added to the clustering. This feature has the least correlation with the noise but at the same time it has very low correlation with other variable too. This variable cannot dominate the clustering procedure. It can be seen also, that the correlation between CHV and tortuosity (0.8871) is almost equal to

the correlation between CHV and noise (0.8693). The numbers of cluster are 7. This is one cluster more than the rank of the matrix. Here is when visual judgment could be a applied to adjust the number of clusters. One realizations seems to be far away in terms of most features (see corresponding figures), and is being clustered alone truly. Yet, there could be features among the feature vector representative that for example the existence

Table 1: Some correlation coefficients for the realizations features.

Correlation	CHV	Net Vol.	Eff. Perm.	noise
CHV	1.000	0.990	0.991	0.851
Net Vol.	0.990	1.000	0.998	0.860
Eff. Perm.	0.991	0.998	1.000	0.860
noise	0.851	0.860	0.860	1.000

Table 2: Some correlation coefficients for the realizations features at the presence of noise.

Correlation	CHV	Net Vol.	Eff. Perm.	Tortuosity	noise
CHV	1.000	0.990	0.991	0.887	0.869
Net Vol.	0.990	1.000	0.998	0.914	0.880
Eff. Perm.	0.991	0.998	1.000	0.903	0.880
Tortuosity	0.887	0.914	0.903	1.000	0.787
noise	0.869	0.880	0.880	0.787	1.000

of two of them is very dominate in clustering. This is very much depending on the selected features. A more variety of the features from different space and areas could bring more variability to the clustering paradigm. The important thing is to be able to consider different aspects when realization selection is performed.

In our second experiment, the total net volume is about 60%. The same clustering algorithm has been applied and similar features have been used. The overall correlation of features are not as large as the first experiment. Seven clusters have been considered and the average clustering error remains the same before and after the addition of noise (E $= 9.6410^{-7}$). Figures 4 and 5 shows less structure compare to the previous experiment. The effect of the variables such as toruosity or number of geoobjects vanishes much faster and a few firs variables seem to control the main part of the clustering task.

Discussions and Results

Applying clustering on the realizations requires more effort than ranking based on a scalar. Depending on the reservoir distribution, the spatial correlation of realizations, the existence heterogeneity, the measured variables, the approach, understandings, decision and results might be different. For example, sometimes the realizations are quite homogenous and many variables are so correlated that they are basically redundant in the clustering analysis. Also, sometimes some variables are too uncorrelated or carry no specific structure so that they would be treated as noise in clustering. Or even if they are able to control clustering, the groups seem more random than conducting an important information of the realizations. In some other cases, features such as CHV seems to dominate the clustering performance. One immediate reason is that CHV measurements is directly related to connected net distribution of deposit. The other reason is also that this relationship has nonlinear structure as it considers factors such as well placement and fluid flow. Also, from practical point of view, this becomes advantageous, since the realization selection mostly takes place for the purpose of flow simulations. In fact, ranking mostly utilize CHV to rank the realizations on clustering.

In our clustering approach, we can specify how important each variable should be in the final grouping. For example, for a case when different CHV for different scales are present and the experiment shows that the CHV of the entire area is mostly dominant in clustering, one could increase the effect of the other smaller scales CHV by squaring (^{n}) them. This all explains that clustering of realizations is not as effective if it just applies automatically. The understating of attributes, visual analysis and one's judgment is fully appreciated.

We have applied Kernel K-means clustering on realizations as a different paradigm in selecting and understanding the uncertainty. We do the clustering over several measured features which together represent every cluster. Different clustering type perform differently and depending on the features' structure one

would perform better than others. In the case of realizations, the measured variables would vary depending on different data analysis or available data. We have found Kernel K-means clustering the most adequate clustering type in our experiment. The reason is that, in Kernel K-means, the partitioning of data takes place in feature space rather than data space. This already resolve the linear partitioning issue with K-means. The nonlinearity of data is captured and transferred to feature space before the linear partitioning is applied to it. In our experiment, CHV dominates the clustering most time due to its particular structure and extra information which carries related to the realizations.

References

G. Tzortzis and A. Likas. The global kernel k-means clustering algorithm. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 1977--1984. IEEE, 2008.

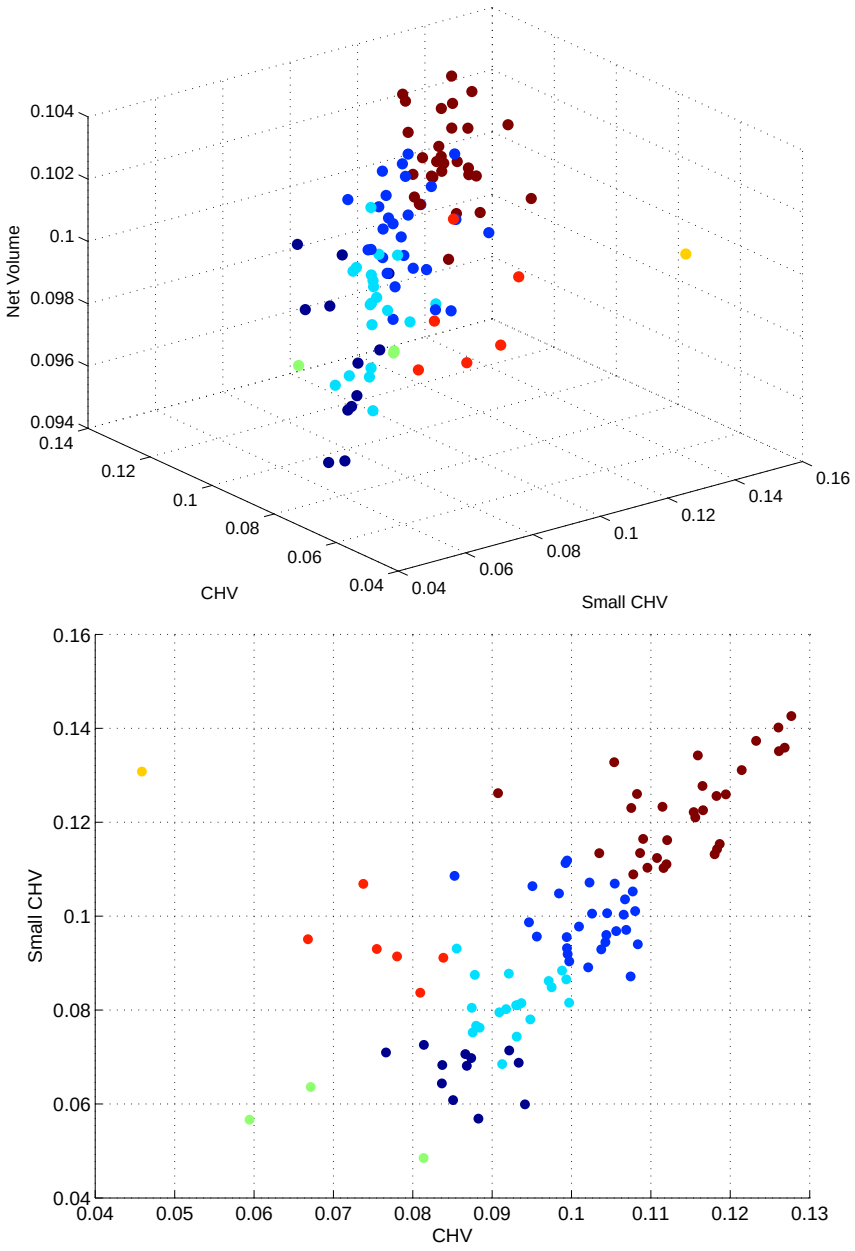

Figure 1: Figure at top illustrates the clustered data in original domain using Kernel K-means clustering. Figure at bottom is the clustered scatter plot of CHV and small CHV.

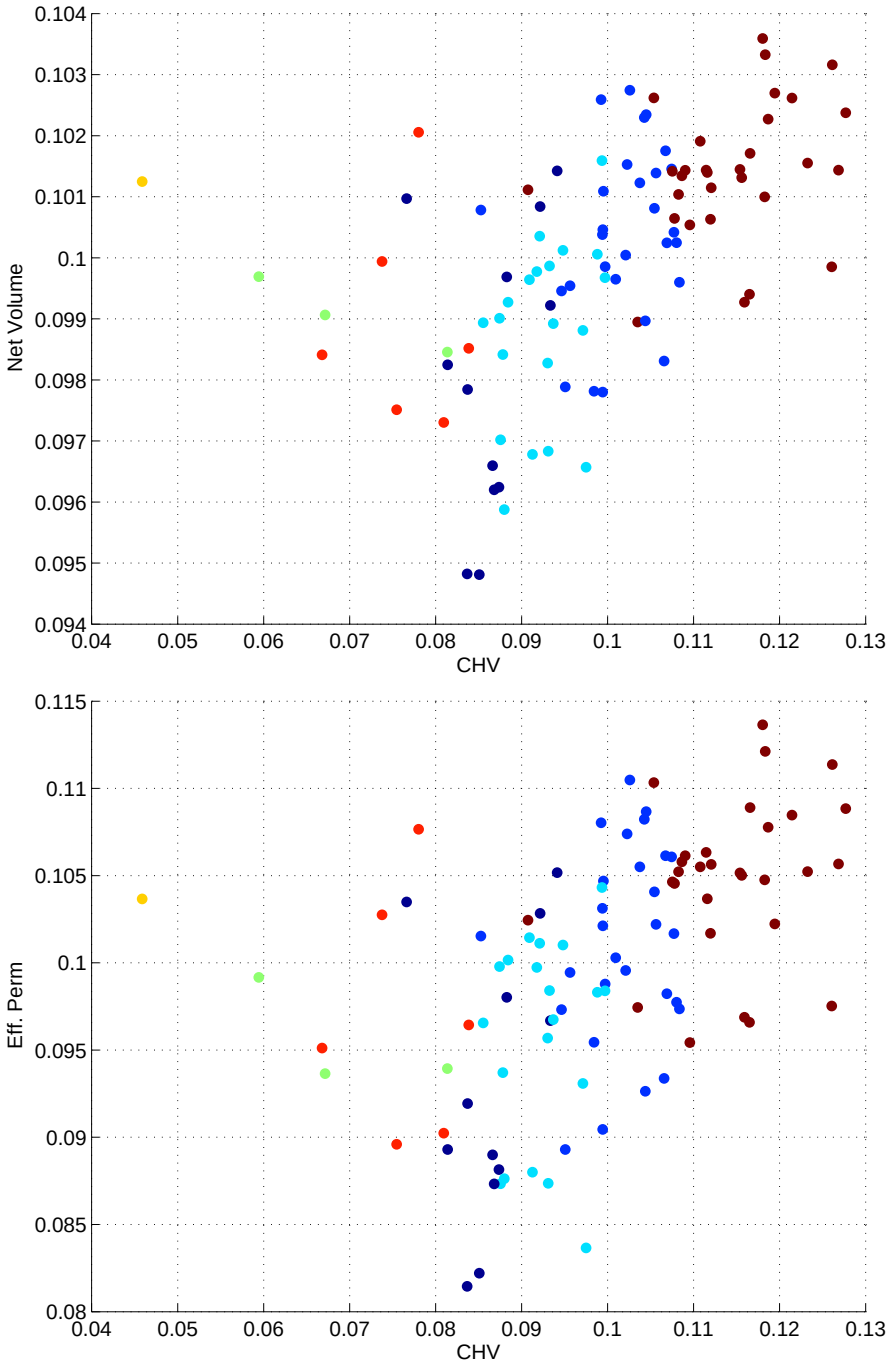

Figure 2: Figure at top illustrates the clustered scatter plot of CHV and net volume. Figure at bottom is the clustered scatter plot of CHV and effective permeability.

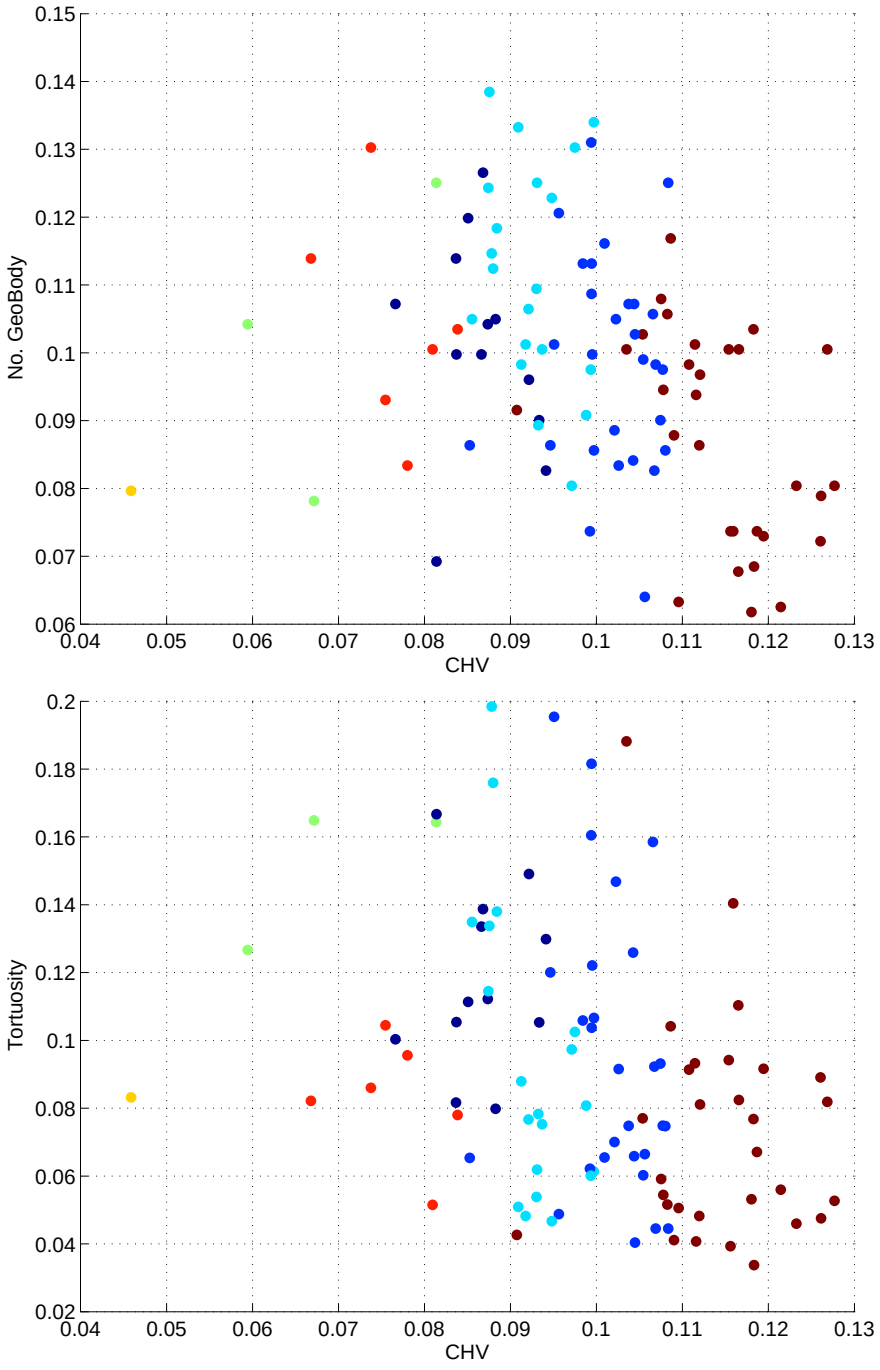

Figure 3: Figure at top illustrates the clustered scatter plot of CHV and numbers of geobodies. Figure at bottom is the clustered scatter plot of CHV and tortuosity.

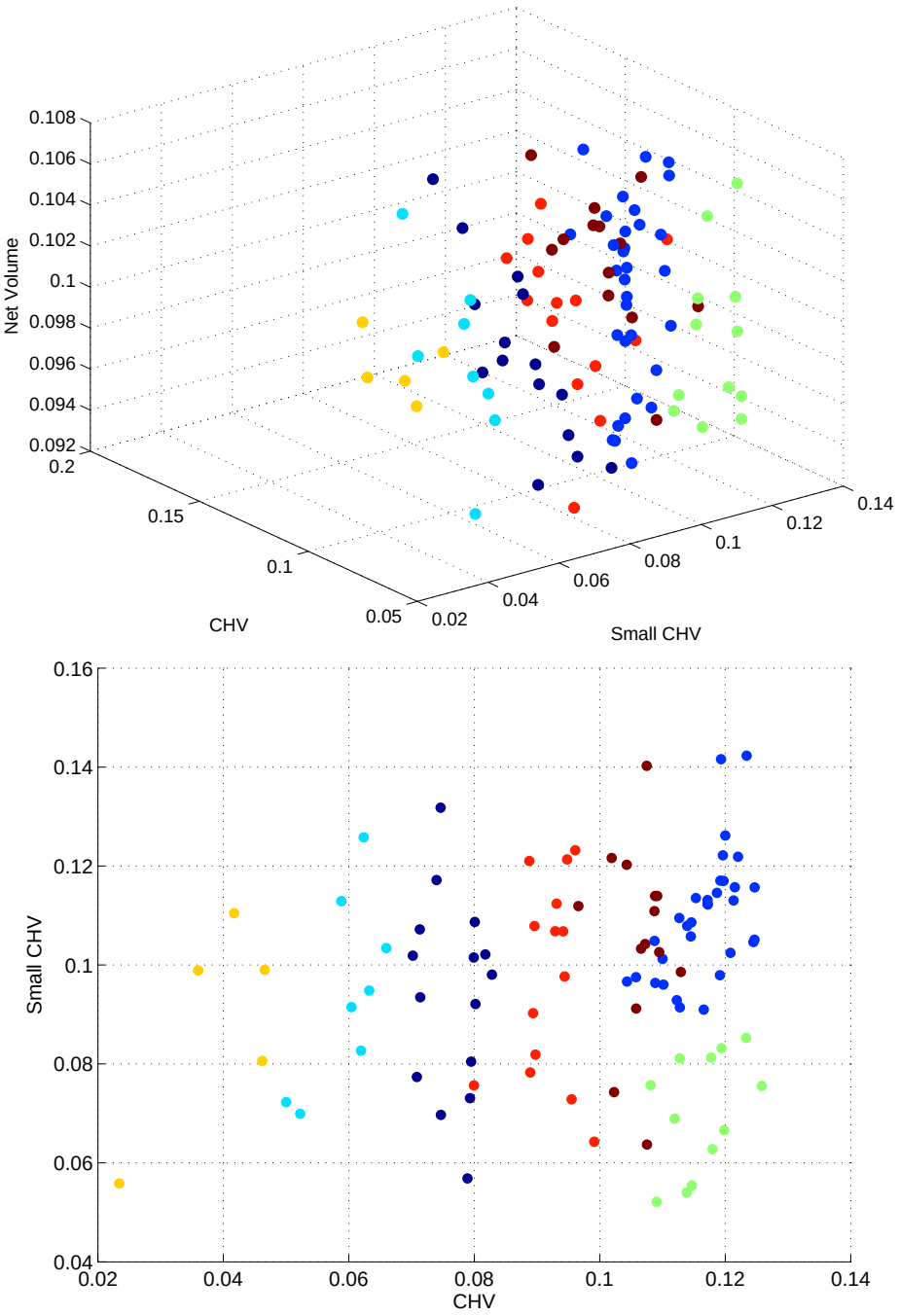

Figure 4: Figure at top illustrates the clustered data in original domain using Kernel K-means clustering. Figure at bottom is the clustered scatter plot of CHV and small CHV.

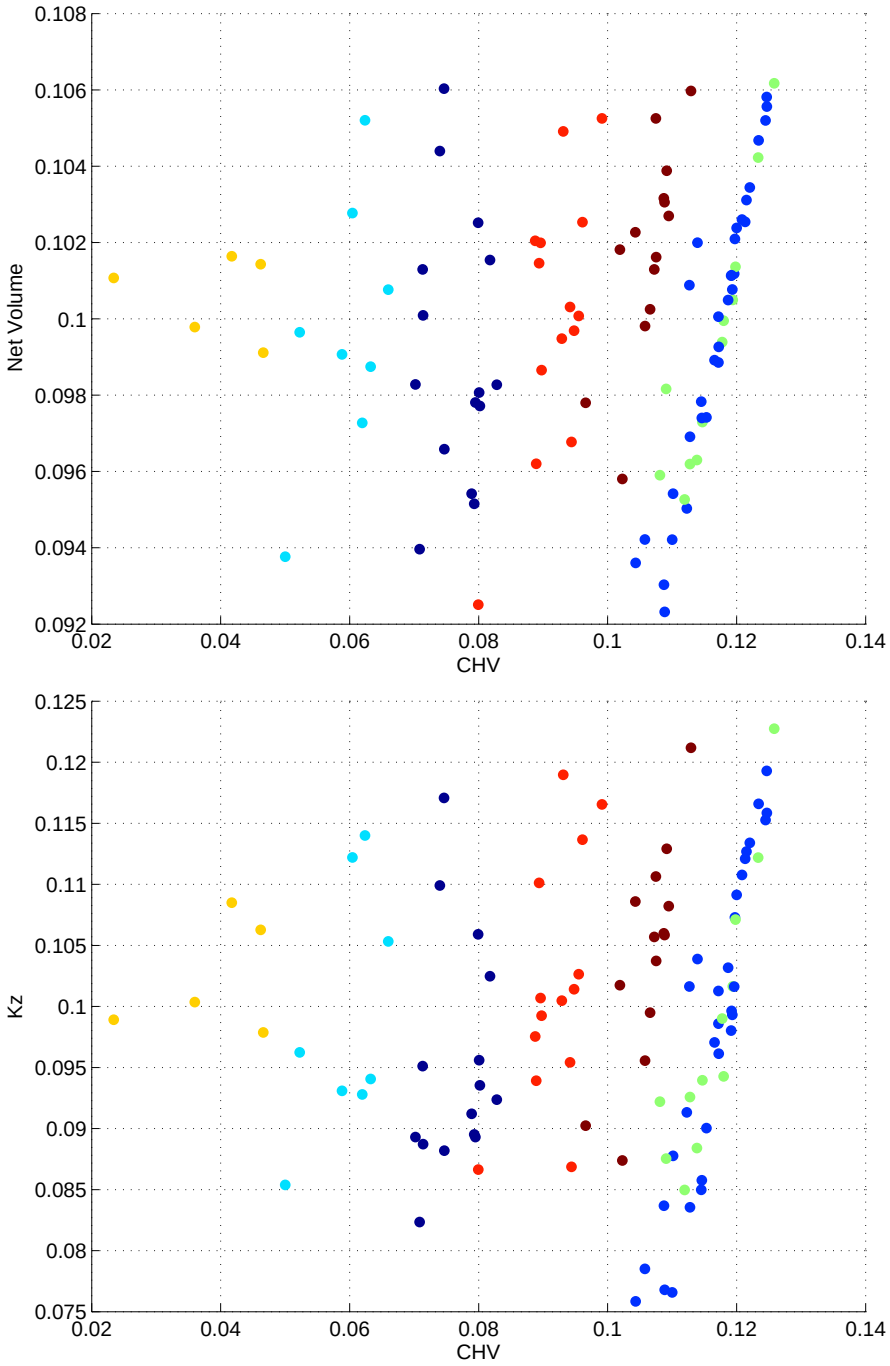

Figure 5: Figure at top illustrates the clustered scatter plot of CHV and net volume. Figure at bottom is the clustered scatter plot of CHV and effective permeability.