Semi-Automated Multiphase Segmentation of 4-D Micro-Computed Tomography Data of Porous Media

Ramaprasad Kulkarni[1], Jeffrey J. Rodriguez[1], and Markus Tuller[2]

THE UNIVERSITY OF ARIZONA

[1] University of Arizona, Dept. of Electrical and Computer Engineering, Tucson, AZ

[2] University of Arizona, Dept. of Soil, Water and Environmental Science, Tucson, AZ

Introduction

- **The development and evaluation of advanced segmentation methods for transformation of grayscale X-ray CT data into a discrete form that allows accurate separation of solid, liquid, and vapor phases for quantitative description of porous media properties and for modeling of dynamic system processes remains a grand challenge.**
- **To advance X-ray CT data segmentation and reduce operator bias, we propose a new semi-automated three-dimensional multiphase algorithm that combines K-means seeding with a Markov random field framework (KM-MRF).**
- **The proposed algorithm was evaluated for two 4-D datasets (Fig. 1), each representing a time-series of 3-D image data, and compared to simple K-means clustering.**

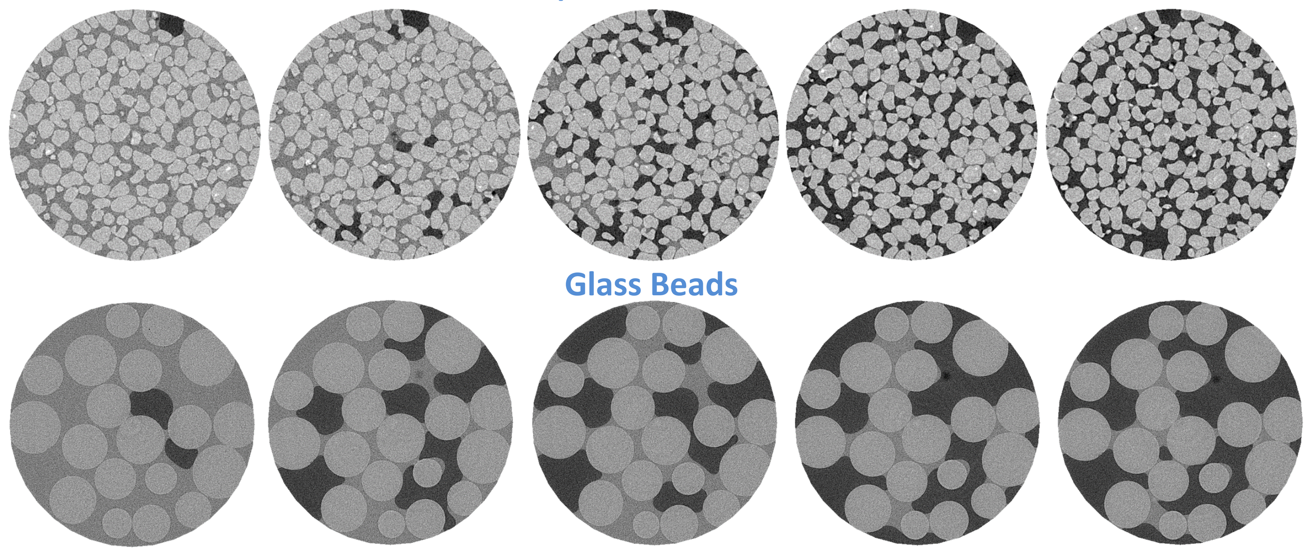

Figure 1: Sample cross sections of raw X-ray CT grayscale data for investigated silica sand and glass beads at various liquid saturation levels.

Markov Random Field (MRF) Segmentation

- **The MRF segmentation objective is to assign a label, representing a phase, to each voxel with a particular gray level proportional to the X-ray attenuation coefficient. Let X represent the set of observed gray levels in the 3-D X-ray CT dataset, and L represent the set of labels of the phase to which each voxel belongs. Applying Bayes' theorem, we can express the posterior probability $P(L/X)$ in terms of likelihood, prior and marginal probabilities** (Kulkarni et al., 2012):

$$P(L\,|\,X) = P(X\,|\,L) * P(L) / P(X)$$

where the likelihood probability $P(X/L)$ represents the probability of gray levels belonging to a particular phase, the prior probability $P(L)$ is the probability distribution of each phase and the marginal likelihood probability $P(X)$ is the probability distribution of observed data.

- **The likelihood probability distribution is assumed to be a Gaussian function and the prior probability is given by the Markov model. The posterior probability can thus be maximized by solving the following objective function** (Kulkarni et al., 2012):

$$\hat{L} = \arg\min_{L} \left(\sum_{i=1}^{N} \left(\ln\sqrt{2\pi\sigma_L} + \frac{(x_i - \mu_L)^2}{2\sigma_L^{\ 2}} \right) + \sum_{\{S_i,S_j\}\in C} \beta * \gamma(l_{S_i}, l_{S_j}) \right)$$

$$\gamma(l_{S_i}, l_{S_j}) = \begin{cases} -1 & if \quad l_{S_i} = l_{S_j} \\ +1 & if \quad l_{S_i} \neq l_{S_j} \end{cases}$$

with μ_L as the mean and σ_L as the standard deviation of labeling L for each phase, l_{s_i} and l_{s_j} as labels for sites S_i and S_j, respectively, corresponding to voxels x_i and x_j in a three dimensional space, and β is a constant that represents the homogeneity between phases.

- **Solving the objective function is a combinatorial optimization problem. We use the deterministic iterated conditional modes algorithm as it yields a local optimum and is computationally less intensive than heuristic algorithms** (Kulkarni et al., 2012).

K-means Seeding

- **The MRF model is inherently powerful for image segmentation because it can generally handle any number of voxel classes (e.g., representing different pore-filling fluids or different solid grain materials). However, it must be initialized with reasonable statistics (i.e., mean and standard deviation) for each voxel class.**
- **We applied K-means (KM) clustering to compute the statistics for each voxel class from the original dataset. The objective of the KM algorithm is to assign the N voxel gray levels, $x_j (1 \leq j \leq N)$ to one of the K clusters, $P_i (1 \leq i \leq K)$ such that the within cluster sum of squares** (WCSS) **is minimized** (Steinley, 2006):

$$\arg\min_{P} \left\{ \sum_{i=1}^{K} \sum_{x_j \in P_i} \left\| x_j - \bar{x}_i \right\| \right\}$$

with $\bar{x}_i$ as the mean of the voxels in the cluster P_i, and $\|\ \ \|$ representing Euclidean distance.

Results

- **Initially saturated samples for silica sand and glass beads were scanned with the synchrotron micro X-ray computed tomography system at the GeoSoilEnviro Consortium for Advanced Radiation Sources (GSECARS) at Argonne National Laboratory** *(data courtesy of Drs. Mark Brusseau and Juliana Araujo Lewis, University of Arizona)*. **After each scan, a small amount of liquid was withdrawn with a syringe and the sample was rescanned to generate 4-D time series data encompassing eight saturation levels for each sample.**
- **First, we only applied K-means clustering to visually evaluate the quality of clustering results. As shown in** Fig. 2, **KM clustering is prone to voxel misclassification. We observed numerous classification errors for both the glass beads and silica sand, which indicates that while KM clustering may perform reasonable well for generating seeding statistics, a more sophisticated locally-adaptive method such as MRF is needed for more accurate segmentation.**

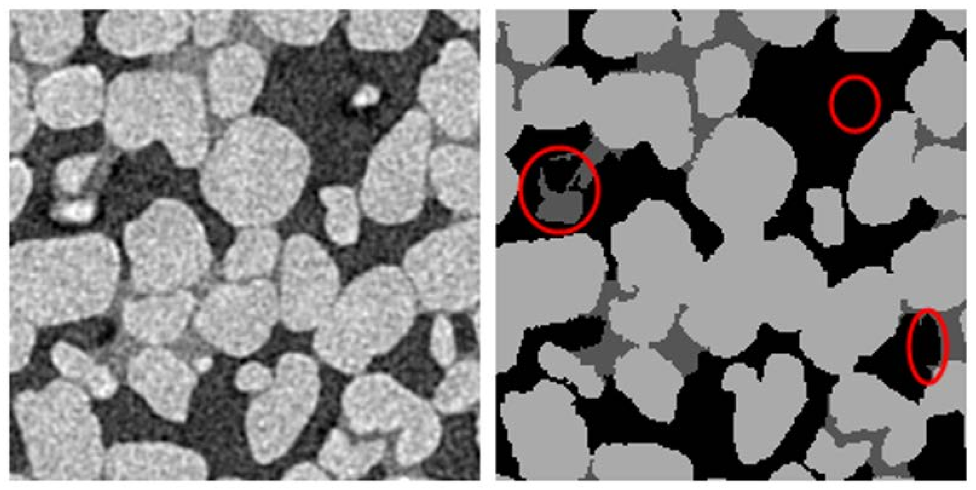

Figure 2: Example for misclassification errors of K-means clustering. Raw grayscale data (left) compared with K-means segmentation results (right).

- **The MRF algorithm with KM seeding performs significantly better** (Fig. 3).

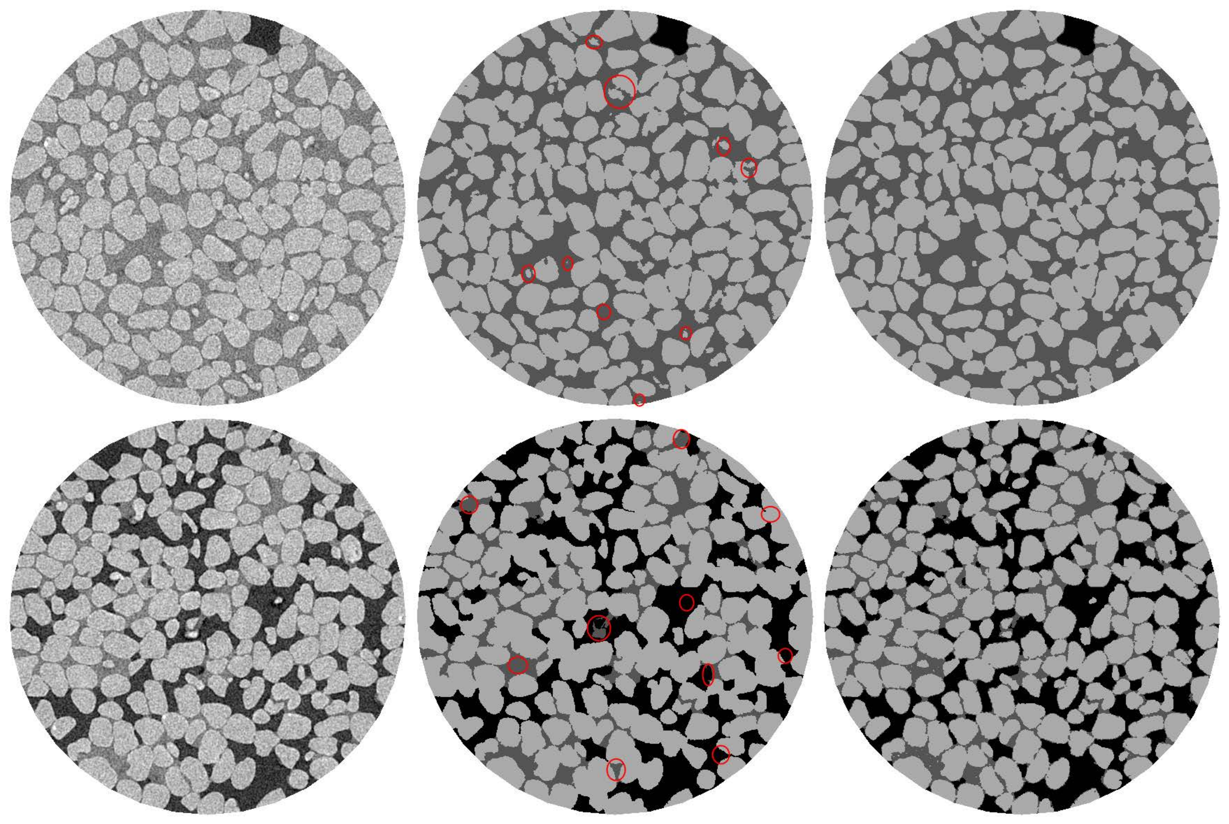

Figure 3: Comparison of KM (middle column) and KM-MRF (right column) segmentation. Raw data are in the left column.

Results Continued

- **The actual porosity for both samples was determined from the container volume, the dry sample mass, and the particle density. CT-derived porosity and associated relative percentage error for the 45/50 silica sand time series dataset is shown in** Fig. 4. **The relative porosity error is computed as $E_A = (\phi_{CT} - \phi_M)/\phi_M$, with ϕ_{CT} as the porosity derived from segmented data, and ϕ_M as the actually measured porosity.**

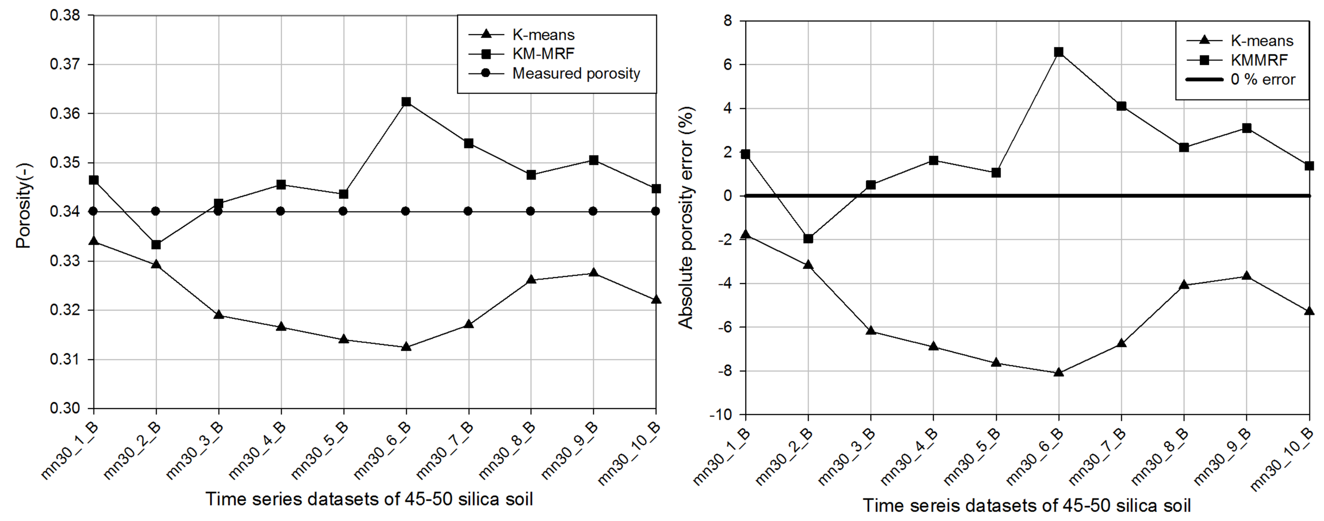

Figure 4: Comparison of actually measured porosity with porosities derived from segmented CT time series data for 45/50 silica sand (left). Relative porosity errors for the same time series dataset (right).

- **The calculated relative porosity errors for KM-MRF segmentation are significantly lower than for sole KM segmentation for silica sand** (Fig.4). **Similar results were obtained for the glass beads.**

Please note that while many available segmentation codes apply 2-D *"slice-by-slice"* processing, all applied algorithms used for this study were coded for true 3-D segmentation, considering the voxel neighborhood in z-direction. Furthermore, the seeding statistics for MRF segmentation were determined once for each data time series for the samples with approximately 50% liquid saturation and then uniformly applied to the remaining samples.

Conclusions and Ongoing Work

- **The introduced semi-automated, two-step KM-MRF algorithm for multiphase segmentation only requires two input parameters, namely the number of phases and the MRF β parameter, which brings us a step closer to the desired unsupervised segmentation, free of operator bias.**
- **Results obtained for time series data for silica sand and glass beads show that the locally adaptive KM-MRF algorithm outperforms the global K-means algorithm.**
- **Potential further improvements could be achieved by considering mixture models (e.g., Gaussian mixture model) and expectation maximization to determine the mean and variance of each phase directly from the 3-D gray level data.**

References

Kulkarni, R., M. Tuller, W. Fink, and D. Wildenschild, 2012. Three-Dimensional Multiphase Segmentation of X-Ray CT Data of Porous Materials Using a Bayesian Markov Random Field Framework. Vadose Zone J., 11, doi:10.2136/vzj2011.0082.

Steinley, D., 2006. K-means clustering: A half-century synthesis. British Journal of Mathematical and Statistical Psychology, 59(1), pp.1-34.

Acknowledgements

Special thanks go to *Drs. Mark Brusseau* and *Juliana Araujo Lewis* from the University of Arizona Department of Soil, Water and Environmental Science for the silica sand and glass beads X-ray CT raw data.