AGU: H13E-1591

MRF-Based Stochastic Joint Inversion of Hydrological and Geophysical Datasets to Characterize Aquifer Heterogeneities

Erasmus K. Oware

Department of Geology, University at Buffalo, SUNY.

AGU: H13E-1591

ABSTRACT

A new stochastic imaging (SI) algorithm is presented. The approach applies Markov random field (MRF) modeling to model aquifer heterogeneities in conformance to a specified site-specific, spatial statistical constraints while honoring the hydrogeophysical measurements. An adaptive algorithm that implicitly infers the applied statistical structure from hydrogeophysical measurements is also presented. Here, the lithological structure of the aquifer and the hydraulic properties within the identified lithologies are estimated simultaneously.

The algorithm is illustrated with hypothetical solute transport experiments with concentration and electrical resistivity monitoring in a heterogeneities binary hydrofacies aquifer. High reconstruction accuracy rates based on the inferred statistics with minimal data conditioning are reported. The algorithm provides a unique potential to improve the computational efficiency of large-scale aquifer characterization problems.

INTRODUCTION

- Stochastic imaging (SI) provides geologically realistic probable outcomes, which is appealing due to our typically limited noisy measurements coupled with our incomplete understanding of subsurface processes.
- Bayesian Markov chain Monte Carlo (McMC) with sequential geostatistical resampling (SGR) algorithms are becoming increasingly popular [*Ruggeri et al*., 2015] These SGR models are variogram- or training image (TI)- driven.

THE PROBLEM

Variograms are based on two-point statistics, which limit their ability to model complex, continuous features [e.g., *Strebelle*, 2002], whereas TI approach to sampling higher-order statistics may bias outputs if the TI's are unrepresentative of the desired process [e.g., *Journel and Zhang*, 2006].

THE PROPOSED SCHEME

- A data-driven (TI-free) SI technique based on Markov random field (MRF) modeling is proposed. MRF modeling is widely used in image processing (e.g., *Geman and Geman*, 1984; *Li*, 2009] and medical imaging [e.g., *Li et al.*, 1995].
- It leverages the equivalence of Gibbs (or Boltzmann) distribution and MRF to identify probable local configurations of an RF in terms of Gibbs energy.

OVERVIEW OF MRF MODELING

MRF modeling is based on a neighborhood system ($\aleph$) within a pixelated RF and the cliques within $\aleph$ [e.g., *Li*, 2009]. Figs. 1 and 2 show an example of a neighborhood system and all its ten associated cliques, respectively.

2	1	2
1	**u**	1
2	1	2

Figure 1: A second order neighborhood system, $\aleph$, (template).

- A subset of $\aleph$ defines a clique if every pair of distinct pixels in the subset are mutual neighbors.

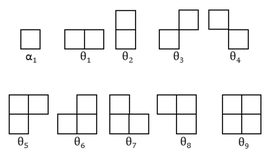

Figure 2: The ten cliques associated with the template in Fig. 1. The α's and θ's represent the parameters of a Gibbs distribution (GD). GD parameters capture spatial statistics, such as size, shape, orientation, clustering, and frequency of regions.

- From the Gibbs-Markov equivalence rule, *Derin and Elliot* (1987) showed that the local conditional distribution of the template is:

$$\Pr[k(u)|\mathbf{k}(\aleph_u)] = \frac{e^{-V(\mathbf{k}(u),\mathbf{k}(\aleph_u),\theta)}}{\Pr[\mathbf{k}(\aleph_u),\theta]} \qquad (1)$$

where $V(\cdot)$ is the clique potential, $\mathbf{k}(u)$ is the value at the central pixel, $\mathbf{k}(\aleph_u)$ are the values at the neighboring sites; and the vector θ contains the GD parameters (spatial statistics) of the cliques (Fig. 2).

- The denominator in Eq. 1 is a constant, hence, *Oware* [in review, WRR] proposes to recast Eq. 1 in a Bayesian inversion framework [*Tarantola*, 2005] as follows:

$$\sigma(\mathbf{k}) = c\,\rho(\mathbf{k})e^{-V(k(u),\mathbf{k}(\aleph_u),\theta)}L(\mathbf{D}_{head},\mathbf{D}_{conc},\mathbf{D}_{geoph}|\mathbf{k}), \qquad (2)$$

where $\sigma(\cdot)$ and $\rho(\cdot)$ represent the posterior and prior distributions, respectively; c is a normalization constant; $e^{-V(k(u),\mathbf{k}(\aleph_u),\theta)}$ assesses the likelihood of $\mathbf{k}(u)$ given $\mathbf{k}(\aleph_u)$ and θ; and $L(\mathbf{D}_{head},\mathbf{D}_{conc},\mathbf{D}_{geoph}|\mathbf{k})$ is the joint likelihood of hydrogeophysical measurements given $\mathbf{k}$.

Eq. 2 is at the heart of the proposed algorithm!

METHODOLOGY

The proposed algorithm proceeds in seven major steps as summarized in the flow chart below.

The primary simulation estimates the lithologic structure conditioned on GD parameters and neighboring values, whereas the secondary simulation samples from representative K distributions of the identified lithologies conditioned on hydrogeophysical measurements.

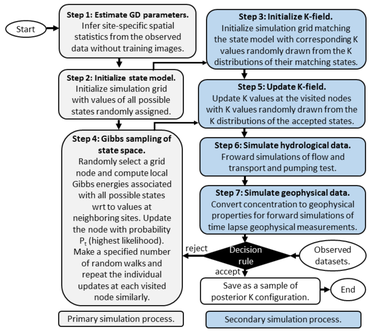

Figure 3: Flow chart of the computational algorithm.

The proposed data-driven adaptive algorithm to implicitly infer GD parameters from Eq. 2 (i.e., step 1 in Fig. 3).

The adaptive algorithm follows the same routine as outlined in Fig. 3, however, the GD parameters here are adaptively resampled for their reconstructions to fit the conditioning dataset.

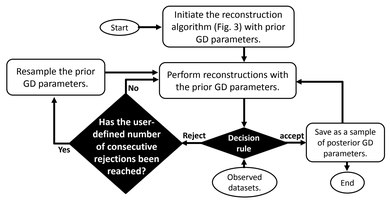

Figure 4: Flow chart of the adaptive algorithm to estimate site-specific GD parameters from conditioning datasets.

TEST CASE: A cross-well synthetic tracer and electrical resistivity (ER) monitoring experiments.

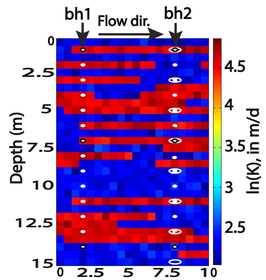

Figure 5: The true K-field with schematic illustration of experimental setup. White filled circles denote locations of ER electrodes (14 in each well), whereas opened ovals represent locations of multilevel concentration sampling ports (eight) in bh2.

- Tracer tests with tracer injection in borehole #1 were simulated with a MATLAB-based flow and transport simulator, MatFlot [*Kunze and Lunati*, 2012].
- Time-lapse concentrations were converted into ER snapshots using Archie's law [*Archie*, 1942].
- A total of 4490 quadrupole measurements were acquired for each time-step using MATLAB-based resistivity forward simulator, FW2_5D [*Pidlisecky and Knight*, 2008].
- Concentrations, ER measurements, and petrophysical conversions were corrupted with white Gaussian noise with standard deviations proportional to 5% of the data values.

PRESENTED WORK

- GD parameters are estimated with the proposed adaptive algorithm.
- Reconstructions are performed with the inferred GD parameters with increasing amount of data conditioning.
- Posterior distributions are estimated based on 15,000 samples.

RESULTS AND DISCUSSION

Estimation of GD parameters based on proposed data-driven (joint concentration and ER) adaptive algorithm (Fig. 4).

- Means of post burn-in samples of 2.3, 1.1, -0.6, and -0.4 (Fig. 6) were estimated for θ_1, θ_2, θ_3 and, θ_4 (see Fig. 2 for their matching cliques), respectively.
- The horizontal GD parameter (θ_1) witnessed the biggest increase from its initial value, which is consistent with the lateral trending of patterns in the target (Fig. 5).
- The estimated GD parameters were applied as calibrated site-specific spatial statistics in all the reconstructions that follow.

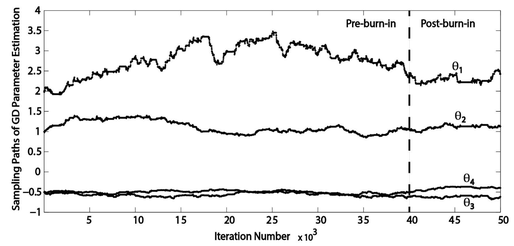

Figure 6: Sampling paths of GD parameter estimation for 50,000 iterations.

Reconstruction based on estimated GD parameters and conditioned on only borehole facies values.

- High reconstruction accuracy rates with limited data conditioning were achieved, with mean facies identification accuracy rate (FIdAR) of 87.7% (Fig. 7).
- The estimated statistics were able to generate patterns that mimic those found in the target, enhancing starting FIdAR of 57% (I) to a maximum FIdAR of 93.2% (III).
- The reconstructions burned-in rapidly (only 29 s to complete 500 iterations).

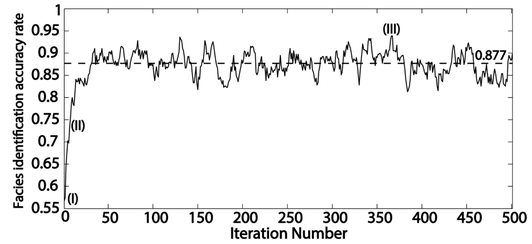

Figure 7: **:** Evolution of facies identification accuracy rates (FIdAR) for 500 iterations. The broken line denotes the mean FIdAR of all realizations after the 30th iteration. Inserted roman numerals mark milestones in the evolution of FIdARs: starting model (I), intermediate (II), maximum FIdAR (III).

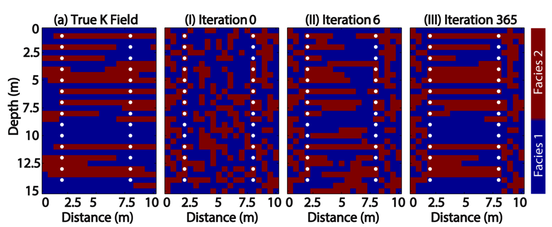

Figure 8: Images associated with the milestones marked in Fig. 7.

Reconstruction based on estimated GD parameters and conditioned on borehole facies values and hydrogeophysical measurements.

- Fig. 9 reveals high estimation accuracy rates of the K-field, which demonstrate the ability of the algorithm to characterize both lithologic and hydraulic properties of an aquifer.
- The inversions required between 3-5 hours to complete 15,000 iterations.

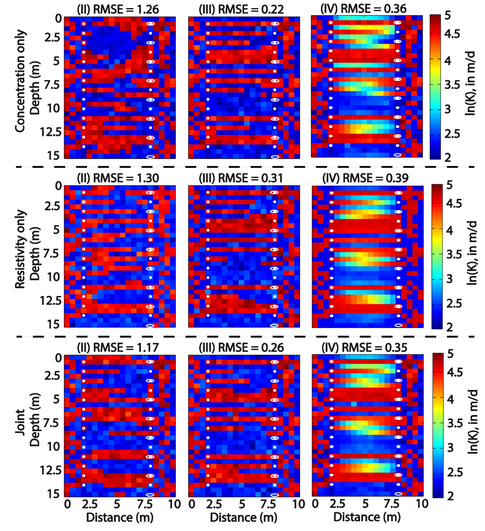

Figure 9: Tomograms of log (K) associated with the milestones marked in the evolution of FIdARs for the hydrogeophysical conditioning (Not shown, similar to Fig. 7). Tomograms conditioned on concentration only (row 1), ER only (row 2), and joint concentration and ER (row 3). K fields associated with: intermediate FIdAR (II), maximum FIdAR (III), and posterior mean K field (IV).

Marginal posterior probability of being facies 2 estimated from the K realizations.

- Results of the posterior probability of being facies 2 (Fig. 10) reveal that the hydraulically conducive zones were accurately identified with high FIdARs of 94.1% (b), 92.6% (c), and 94.4% (d).

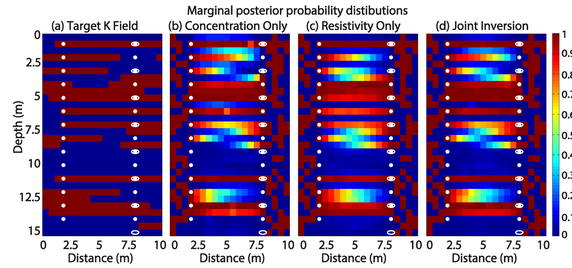

Figure 10: Images of marginal posterior probability of being facies 2. Target model (a); inversion results conditioned on: concentration only (b), resistivity only (c), and joint concentration and ER (d).

CONCLUSIONS

- Potential application of MRF modeling to characterize aquifer heterogeneities constrained to site-specific spatial statistics and joint hydrogeophysical measurements has been demonstrated.
- A key finding is the high reconstruction accuracy rates obtained from the inferred statistics with limited data conditioning in a computationally efficient manner. This implies that site-specific statistics can be captured at a high resolution cross-borehole scale and the calibrated statistics applied for "watershed-scale" reconstruction conditioned on low-resolution data at a reduced computational cost.
- The algorithm also presents a viable technique to characterize discrete fracture networks (DFN), i.e., site-specific fracture statistics, such as fracture size, spacing, orientation, and density can be captured in the form of GD parameters, and subsequently applied to reconstruct DFN representative of the site.

References

- Archie, G.E. (1942), The electrical resistivity log as an aid in determining some reservoir characteristics, *Trans. Am. Inst. Min. Metall. Pet. Eng*., 146, 54–62.
- Derin, H., and H. Elliott (1987), Modeling and segmentation of noisy and texture images using Gibbs random fields, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, pp. 39-55.
- Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images." *IEEE trans. Pattern Analysis and Machine intelligence,* vol. 6, pp. 721-741.
- Irving, J. D., and K. Singha (2010), Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *WRR*, 46, no. 11, W11514.
- Journel, A., and T. Zhang (2006), The necessity of a multiple-point prior model.*Mathematical geology*, 38(5), 591-610.
- Künze, R., and I. Lunati (2012), An adaptive multiscale method for density-driven instabilities. *Journal of Computational Physics*, 231(17), 5557-5570.
- Li, H. D., M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark (1995), Markov random field for tumor detection in digital mammography. *Medical Imaging, IEEE Transactions on*, *14*(3), 565-576.
- Li, S. Z. (2009), Markov random field modeling in image analysis, Springer Science & Business Media.
- Pidlisecky, A., and R.J. Knight (2008), FW2_5D: A MATLAB 2.5-D electrical resistivity modeling code. *Computers & Geosciences*, 34(12), 1645-1654.
- Tarantola, A. (2005), Inverse problem theory and methods for model parameter estimation, *Society of Industrial and Applied Mathematics*.