



## Aquifer structure identification using stochastic inversion

Dylan R. Harp,<sup>1,2</sup> Zhenxue Dai,<sup>2</sup> Andrew V. Wolfsberg,<sup>2</sup> Jasper A. Vrugt,<sup>3</sup>  
Bruce A. Robinson,<sup>4</sup> and Velimir V. Vesselinov<sup>2</sup>

Received 11 February 2008; revised 13 March 2008; accepted 17 March 2008; published 24 April 2008.

[1] This study presents a stochastic inverse method for aquifer structure identification using sparse geophysical and hydraulic response data. The method is based on updating structure parameters from a transition probability model to iteratively modify the aquifer structure and parameter zonation. The method is extended to the adaptive parameterization of facies hydraulic parameters by including these parameters as optimization variables. The stochastic nature of the statistical structure parameters leads to nonconvex objective functions. A multi-method genetically adaptive evolutionary approach (AMALGAM-SO) was selected to perform the inversion given its search capabilities. Results are obtained as a probabilistic assessment of facies distribution based on indicator cokriging simulation of the optimized structural parameters. The method is illustrated by estimating the structure and facies hydraulic parameters of a synthetic example with a transient hydraulic response. **Citation:** Harp, D. R., Z. Dai, A. V. Wolfsberg, J. A. Vrugt, B. A. Robinson, and V. V. Vesselinov (2008), Aquifer structure identification using stochastic inversion, *Geophys. Res. Lett.*, 35, L08404, doi:10.1029/2008GL033585.

### 1. Introduction

[2] Determining field-scale parameters in sufficient detail to capture aquifer heterogeneities is one of the greatest challenges for predicting flow and contaminant transport in large-scale subsurface systems. Uncertainty in the form of conceptual model bias can be introduced in modeling groundwater flow and contaminant transport when aquifer structure is fixed based on sparse or incomplete geophysical data [Chen and Rubin, 2003]. Once the structure has been fixed in this way, focus is placed on model calibration of zoned parameters through parameter estimation. This type of inverse method has been applied in many instances to estimate flow and transport parameters at various spatial scales [e.g., Cooley, 1983; Carrera and Neuman, 1986; Dai and Samper, 2006]. Deterministic aquifer structure may introduce larger bias and uncertainty into a model than an inappropriate choice of facies hydraulic parameters [Ye et al., 2004; Lu and Robinson, 2006].

[3] Given uncertainty in both aquifer structure and hydrologic parameters Sun [2005] suggested “adaptive parameterization” to couple structure identification and parameter estimation in contaminant transport model calibration. This provides a more complete inversion of the model, allowing the optimization to explore combinations of structure geometry and hydraulic parameters. Previous studies usually assumed that the aquifer parameter zonations are randomly distributed. However, recent geological and geostatistical studies indicate that aquifer facies distributions are spatially correlated [Carle and Fogg, 1997; Ritzi et al., 2004].

[4] We propose a structure identification method that accounts for spatial correlation by means of a stochastic inversion of a transition probability model, in an analytical framework [Dai et al., 2007], describing facies volume proportions, mean lengths, and juxtapositioning. The transition probability model provides a nonparametric, Markov chain approach to indicator geostatistics that is well suited to applications with sparse information [Carle and Fogg, 1997]. The analytical solution allows structure identification to be cast as a conventional inverse modeling problem, using statistical structure parameters (such as facies volume proportions and mean lengths) to iteratively update the transition probability model. The facies proportions and mean lengths define the transition probability matrix. Indicator cokriging simulation produces the aquifer facies distributions, ensuring the statistical properties defined by the transition probability model are maintained. In this way, the aquifer structure is updated in the inversion, while the information provided by the conditional data (the sparse geological and geophysical data used to describe the facies distribution in boreholes) is honored. The optimization of the model inversion is performed using a genetically-adaptive multi-method search algorithm called AMALGAM-SO. This method was chosen as it combines the strengths of several different evolutionary search approaches and has been shown to achieve good efficiency across a range of difficult synthetic benchmark problems [Vrugt et al., 2008]. While other optimization algorithms could potentially be implemented to drive the stochastic inversion, AMALGAM-SO was selected to illustrate the stochastic inversion methodology, without comparing its performance to other optimization algorithms. This decision was partly based on the assumption that, although an analytical solution of the transition probability model is utilized here, gradient-based methods would still have difficulty given the stochastic nature of the structural variables, which serve as inputs to the stochastic simulation. The analytical solution of the transition probability model provides the computational efficiency necessary for the large number of model evaluations required by the stochastic inversion. The combination of the analytical solution of

<sup>1</sup>Department of Civil Engineering, University of New Mexico, Albuquerque, New Mexico, USA.

<sup>2</sup>Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

<sup>3</sup>Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

<sup>4</sup>Civilian Nuclear Programs, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

the transition probability model and AMALGAM-SO provides a robust, computationally efficient model inversion with the ability to deal with complex fitness response surfaces. While in the past, the stochastic inversion described here was not possible, due to computational and algorithmic limitations, we propose that through the use of modern computers and analytically derived structure parameters [Dai *et al.*, 2007], this type of inversion can be realized.

## 2. Facies Transition Probability Model

[5] Transition probability models have been used by geologists to describe sediment facies distributions for a few decades [e.g., Agterberg, 1974; Carle and Fogg, 1997]. Recently, Ritzi *et al.* [2004] and Dai *et al.* [2005] incorporated the work of Carle and Fogg [1997] to relate the structure of the indicator random variables to proportions, geometry, and pattern of the aquifer facies. Under two assumptions: (1) the cross-transition probabilities depend on facies volumetric proportions only, and (2) the juxtapositional tendencies between categories  $k$  and  $j$  are assumed symmetric in the direction  $\phi$ , Dai *et al.* [2007] derived an analytical solution for the transition probability model as

$$t_{ki}(\mathbf{h}_\phi) = p_i + (\delta_{ki} - p_i)e^{-\frac{h_\phi}{\lambda_I}} \quad (k, i = \overline{1, N}), \quad (1)$$

where  $t_{ki}(\mathbf{h}_\phi)$  is the probability of transitioning from facies  $k$  to facies  $i$  in lag distance  $\mathbf{h}$  in direction  $\phi$ ,  $p_i$  is the proportion of facies  $i$ ,  $\delta_{ki}$  is the Kronecker delta,  $\lambda_I$  is the indicator correlation length, and  $N$  is the number of facies. By taking the partial derivative of the auto-transition probability (equation (1) with  $k = i$ ) with respect to  $\mathbf{h}$ , the mean length ( $\bar{L}_{k,\phi}$ ) can be related to  $\lambda_I$  as [Dai *et al.*, 2007]

$$\left. \frac{\partial t_{kk}(\mathbf{h}_\phi)}{\partial \mathbf{h}_\phi} \right|_{\mathbf{h}_\phi \rightarrow 0^+} = -\frac{1}{\lambda_I}(1 - p_k) = -\frac{1}{\bar{L}_{k,\phi}}. \quad (2)$$

Equation (2) defines the relationship between indicator correlation length and the statistical parameters of facies proportion and mean lengths. Using equations (1) and (2), the continuous-lag transition probability matrix  $\mathbf{T}$  in direction  $\phi$  can be defined by the facies proportions and mean lengths as  $\mathbf{T}(\mathbf{h}_\phi) = (t_{ki}(\mathbf{h}_\phi))_{N \times N}$ . The transition probability matrix can be used for indicator simulation of aquifer facies using the indicator cokriging method [Carle and Fogg, 1997].

## 3. Stochastic Inverse Method

[6] The transition probability model establishes a bridge between aquifer statistical parameters and aquifer facies distributions. By estimating these statistical parameters, we are able to formulate a model inversion to identify aquifer structure. A flow diagram of the stochastic inversion method is presented in Figure 1.

[7] The transition probability model is updated by calculating the transition probability matrix, using values of facies lengths and proportions generated as offspring of the previous generation of solutions. The structure is updated by indicator cokriging simulation using the updated transition probability model, where a single reali-

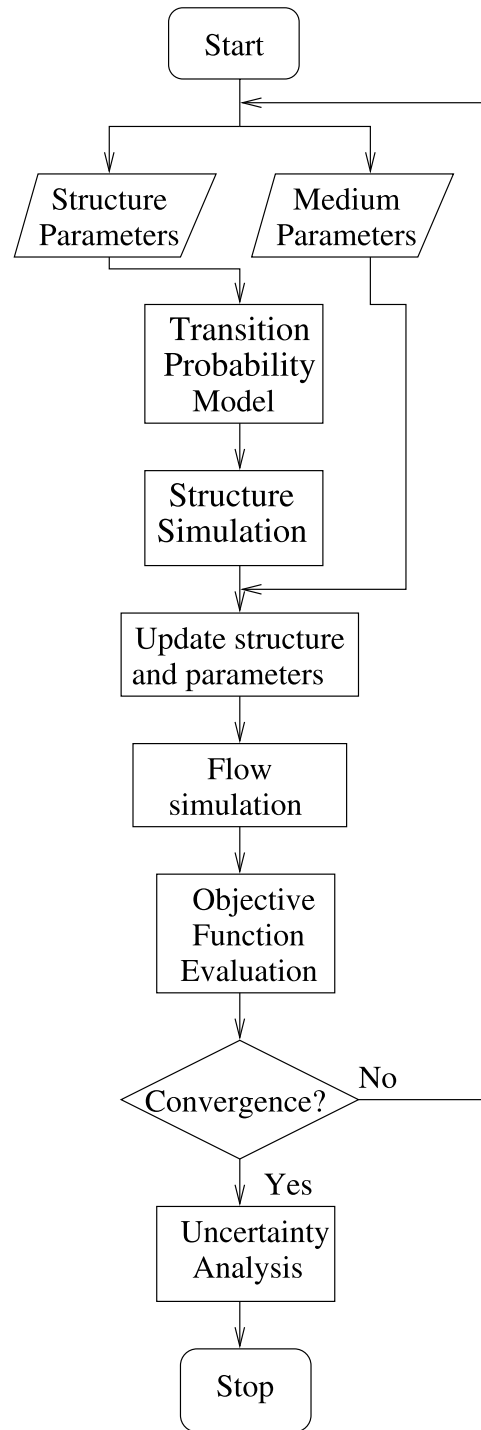


Figure 1. Stochastic inversion flow diagram.

zation is used to represent the collection of equally probable realizations of the given transition probability model. Updated facies hydraulic parameters are applied to the structure zonation. Flow simulation is performed using the Finite Element Heat and Mass transfer (FEHM) code [Zyvoloski *et al.*, 1997], employing observed or assumed flow and boundary conditions, producing simulated transient head data.

[8] The inverse modeling is performed with the goal of minimizing residuals of hydraulic head, where the objective function,  $J$ , can be defined as

$$J = \min_{\beta \in \mathbf{B}} \sum_{i=1}^M (\hat{h}_i(\beta) - h_i)^2, \quad (3)$$

where  $\hat{h}(\beta)$  is an estimated head using parameter values in the vector  $\beta$  constrained by  $\mathbf{B}$ , which is defined by the upper and lower parameter bounds, where  $\mathbf{B} \subseteq \mathbb{R}^p$ ,  $p$  being the number of parameters,  $h$  is a measured head, and  $M$  is the number of measured heads.

[9] Equation (3) defines the fitness function optimized with AMALGAM-SO [Vrugt *et al.*, 2008]. In general, AMALGAM-SO allocates the number of offspring at each population size,  $N^l = \{N_1^l, \dots, N_q^l\}$ , to  $q$  algorithms using a weighting scheme based on each algorithm's previous performance, where  $l$  is the population size index. In this way, AMALGAM-SO is able to exploit the individual strengths of selected algorithms at various stages of the optimization. AMALGAM-SO employs a population incrementing restart strategy as its basis for collecting algorithm performance information used to update the offspring allocation [Vrugt *et al.*, 2008]. This method has the advantage of combining individual algorithm strengths by allowing algorithms to exchange search information, and by adaptively distributing preference to algorithms exhibiting superior performance. In the current study, Covariance Matrix Adaptation (CMA), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) strategies were selected for the  $q = 3$  optimization algorithms, as this combination has shown improved performance over other combinations [Vrugt *et al.*, 2008]. The sequence of population sizes used in the current research was  $l = 5, 10, 15, 20$ . For more details on the settings of AMALGAM-SO, refer to [Vrugt *et al.*, 2008].

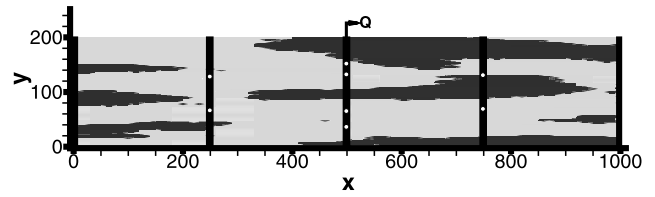
[10] The uncertainty associated with the resulting optimized transition probability model is evaluated using conditional simulation and presented as the final result in the form of a structural probability map, which in the present case defines the shape and location of clay facies within a background of sand. A map of clay probability is produced by approximating the one-location marginal probabilities for clay by the sample mean of the clay indicator spatial function  $I_{clay}(\mathbf{x})$  with respect to the optimized structure parameters as

$$p_{clay}(\mathbf{x}) = Pr\{I_{clay}(\mathbf{x}) = 1\} = \frac{1}{R} \sum_{i=1}^R I_{clay,i}(\mathbf{x}), \quad (4)$$

where  $R$  is the number of realizations.

#### 4. Synthetic Example

[11] The stochastic inversion method is illustrated by means of a synthetic example of a confined aquifer with borehole geophysical data and transient head measurements from a synthetic pump test. The distribution of clay and sand in cross-section, where clay facies are embedded within a background of sand, is illustrated in Figure 2. In this example, the permeability of sand is  $10^{-10} \text{ m}^2$  and the permeability of clay is  $10^{-13} \text{ m}^2$ . The synthetic structure



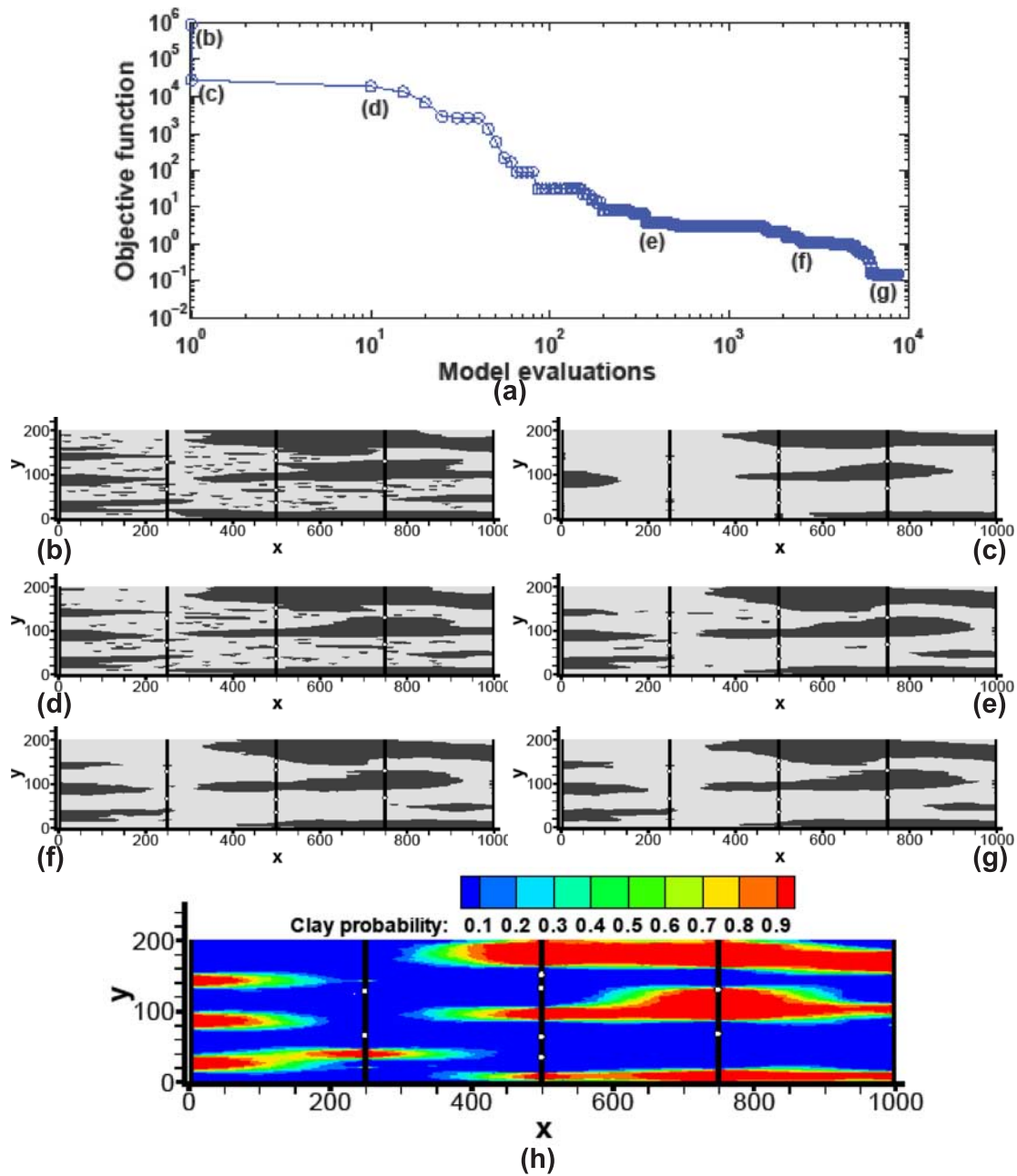
**Figure 2.** Synthetic distribution where light grey represents sand and dark grey represents clay. The black vertical lines indicate observation and pumping wells, as well as locations where facies indicator data has been collected. The white dots indicate head observation locations.

was generated by conditional simulation using prespecified structure parameters. The proportion of sand ( $p_s$ ) in the synthetic example is set at 0.7, while the proportion of clay ( $p_c$ ) is 0.3. The mean lengths of the clay in the  $x$  (length) and  $y$  (thickness) directions are 300 m and 20 m, respectively. Conditional data, which comprises the indicator data of the facies distribution, is collected as continuous bore log data from observation wells defined along the transect at  $x = 0, 250, 500, 750,$  and  $1000$  m indicated by the vertical lines in Figure 2. The well at  $x = 500$  m is set as the pumping well with a constant flow rate of  $Q = 10.2 \text{ kg/s}$ . The boundaries at  $x = 0$  m and  $x = 1000$  m are set as constant head boundaries with heads of 100 m and 95 m, respectively. The top and bottom boundaries at  $y = 0$  m and  $y = 200$  m are no-flow boundaries for this confined model formulation. Measured transient heads were collected at 20 discrete times over a one year simulation, where the size of the time step increased over the simulation. The measurements were collected at 8 locations along the three central wells indicated by white dots in Figure 2, where it has been assumed that the pumping well has the ability to measure pressures while pumping.

[12] This synthetic example was constructed to emulate typical applications encountered in practice, while still allowing a complete evaluation of the robustness and efficiency of the stochastic inversion with respect to a known structure and parameterization.

#### 5. Results and Discussion

[13] Figure 3a presents a plot of the objective function as a function of the number of model evaluations, where it is apparent by the step-like decreases in the objective function that AMALGAM-SO continues to locate improved solutions throughout the inversion. Figures 3b and 3c present the distributions resulting at the lower and upper bounds of the parameters, respectively, defined in the first two rows of Table 1. These two scenarios represent two points along the convex hull of the solution space  $\mathbf{B} \subseteq \mathbb{R}^p$  that the stochastic inversion is required to explore. Inspection of these two plots indicates the diversity of structures considered in the inversion. Figures 3d, 3e, and 3f present distributions at key points during the progression of the inversion, while Figure 3g presents the distribution of the optimal parameters. By inspecting these distributions, the transformation towards the synthetic distribution is apparent. These stages of the optimization are indicated in Figure 3a by their subfigure letter.



**Figure 3.** (a) A plot of the objective function versus the number of model evaluations and (b–g) the aquifer structure as different stages of the stochastic inversion, with their corresponding locations noted in Figure 3a. Refer to Table 1 for detailed information on these aquifer structures. (h) Also shown is the clay probability map produced by stochastic simulation of the optimized structural parameters.

**Table 1.** Structure and Hydraulic Parameters at the Lower (1a) and Upper (1b) Parameter Bounds, at Various Stages of the Model Inversion, and for the Synthetic Example<sup>a</sup>

Figure	Iteration Number	Proportion		Clay Mean Length, m	Clay Mean Thickness, m	Perm. log <sub>e</sub> m <sup>2</sup>		Objective Function
		Sand	Clay			Sand	Clay	
3b	1a	0.6	0.4	280	15	-11	-14	876109
3c	1b	0.8	0.2	320	25	-9	-12	28459
3d	10	0.63	0.37	285.1	15.9	-9.33	-12.31	18440
3e	345	0.67	0.33	310.5	22.3	-9.99	-12.35	3.63
3f	2500	0.69	0.31	309.3	22.0	-10.00	-12.68	1.15
3g	6900	0.69	0.31	319.4	23.2	-10.00	-12.89	0.15
2	True	0.7	0.3	300	20	-10	-13	0

<sup>a</sup>Refer to the figures identified in column 1 for plots of the distributions.

[14] Corresponding tabular information on these distributions, and the synthetic distribution (Figure 2), are presented in Table 1, including iteration number, parameter values, and objective function values. By inspecting the decrease of the objective function plotted in Figure 3a and listed in Table 1, it is apparent that dramatic improvements are made during the course of the optimization, resulting in an extremely small objective function value for the optimized solution. This indicates that an estimate of the aquifer structure has been obtained that closely mimics the hydraulic response of the aquifer. Although an excellent fit to the observed aquifer response has been achieved, it is apparent in Table 1 that there is some discrepancy between the parameters used to generate the synthetic aquifer response, and those estimated with AMALGAM-SO. This is explained by the stochastic nature of this inverse problem. Multiple realizations for the same parameter combination result in widely varying values of the objective function. For instance, when evaluating 100 realizations generated using the optimized parameters, objective function values are obtained that vary between 0.5 and 200, with one outlier around 600. The standard deviation of these objective function values is approximately 68.9. This stochasticity allows different parameter combinations to generate nearly similar responses of the aquifer. The interest is therefore not so much in the exact values of the parameters, but on the optimized structure that has been identified. The latter has been successfully achieved, considering the very small value of the objective function, and the close similarity between the true and inversely estimated facies grid is assigned to the correct facies with respect to the synthetic example.

[15] The final result is presented in a probability map of clay facies distribution in Figure 3h where the estimated one-location marginal probabilities of clay (equation (4)) were calculated from 100 realizations based on the optimized structure parameters. While this does provide the uncertainty of the structure based on the optimized transition probability model, it does not indicate the uncertainty of the structure with respect to the aquifer response. Future work will expand this uncertainty analysis by identifying the set of plausible structures with regard to aquifer response. It is apparent from a comparison of the true distribution (Figure 2) and the resulting clay probability map (Figure 3h) that the large structural features are captured in the analysis. These results demonstrate that the stochastic inversion is able to reduce the objective function to a reasonably low value (indicating that the response of the aquifer is modeled accurately) and that the large structural features of the aquifer are identified adequately given the limited amount of information provided in the synthetic data.

## 6. Conclusions

[16] A method has been developed for aquifer structure identification using stochastic inversion. The method is based on updating structure parameters of a facies transition probability model to identify aquifer structure. This approach allows the problem to be formulated in an inverse modeling framework that can be extended to adaptive

parameterization of facies hydraulic parameters by taking advantage of the efficiency in model evaluations with the analytical solution of the facies transition probability model. The inversion is driven by a multi-method genetically adaptive evolutionary optimization approach [Vrugt et al., 2008], providing a robust inversion capable of traversing the complicated fitness landscape. Results are obtained as a probabilistic estimate of the existence of facies at a particular location given the optimized statistical structure parameters. This method can be applied to pump-test datasets, where some geophysical data are available, to provide probabilistic estimates of the facies distributions based on the hydraulic responses of the aquifer.

[17] **Acknowledgments.** The reported research was supported by Los Alamos National Laboratory's Directed Research and Development Project. The first author is supported as a Graduate Research Assistant at the Los Alamos National Laboratory. The fourth author is supported by a J. Robert Oppenheimer Fellowship from the LANL postdoctoral program. The authors thank Edward Kwicklis, George Zvyoloski, and Zhiming Lu for their constructive comments to the manuscript.

## References

- Agterberg, F. (1974), *Geostatistics*, 596 pp., Elsevier Sci., New York.
- Carle, S., and G. Fogg (1997), Modeling spatial variability with one and multidimensional continuous-lag Markov chains, *Math. Geol.*, 29(7), 891–918.
- Carrera, J., and S. Neuman (1986), Estimation of aquifer parameters under steady-state and transient conditions: 2. Uniqueness, stability, and solution algorithms, *Water Resour. Res.*, 22(2), 211–227.
- Chen, J., and Y. Rubin (2003), An effective Bayesian model for lithofacies estimation using geophysical data, *Water Resour. Res.*, 39(5), 1118, doi:10.1029/2002WR001666.
- Cooley, R. (1983), Incorporation of prior information on parameters into nonlinear regression groundwater flow models: 2. Applications, *Water Resour. Res.*, 19(3), 662–676.
- Dai, Z., and J. Samper (2006), Inverse modeling of water flow and multi-component reactive transport in coastal aquifer systems, *J. Hydrol.*, 327, 447–461.
- Dai, Z., R. W. Ritz Jr., and D. F. Dominic (2005), Improving permeability semivariograms with transition probability models of hierarchical sedimentary architecture derived from outcrop analog studies, *Water Resour. Res.*, 41, W07032, doi:10.1029/2004WR003515.
- Dai, Z., A. Wolfsberg, Z. Lu, and R. Ritz Jr. (2007), Representing aquifer architecture in macrodispersivity models with an analytical solution of the transition probability matrix, *Geophys. Res. Lett.*, 34, L20406, doi:10.1029/2007GL031608.
- Lu, Z., and B. A. Robinson (2006), Parameter identification using the level set method, *Geophys. Res. Lett.*, 33, L06404, doi:10.1029/2005GL025541.
- Ritz, R. W., Z. Dai, D. F. Dominic, and Y. N. Rubin (2004), Spatial correlation of permeability in cross-stratified sediment with hierarchical architecture, *Water Resour. Res.*, 40(3), W03513, doi:10.1029/2003WR002420.
- Sun, N.-Z. (2005), Structure reduction and robust experimental design for distributed parameter identification, *Inverse Problems*, 21, 739–758.
- Vrugt, J., B. Robinson, and J. Hyman (2008), A universal multimethod search strategy for computationally efficient global optimization, *IEEE Trans. Evol. Comput.*, in press.
- Ye, M., S. Neuman, and P. Meyer (2004), Maximum likelihood bayesian averaging of spatial variability models in unsaturated fractured tuff, *Water Resour. Res.*, 40, W05113, doi:10.1029/2003WR002557.
- Zvyoloski, G., B. Robinson, Z. Dash, and L. Trease (1997), Summary of the models and methods for the Fehm application—A finite-element heat- and mass-transfer code, *Tech. Rep. LA-13307-MS*, Los Alamos Natl. Lab., Los Alamos, N. M.

Z. Dai, D. R. Harp, V. V. Vesselinov, and A. V. Wolfsberg, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. (dharp@lanl.gov)

B. A. Robinson, Civilian Nuclear Programs, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

J. A. Vrugt, Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.