COUPLING SEQUENTIAL-SELF CALIBRATION AND GENETIC ALGORITHMS TO INTEGRATE PRODUCTION DATA IN GEOSTATISTICAL RESERVOIR MODELING

XIAN-HUAN WEN, TINA YU and SEONG LEE
Chevrontexaco Energy Technology Company, San Ramon, CA 94583, USA

Abstract. The sequential-self calibration (SSC) method is a geostatistical-based inverse technique that allows fast integration of dynamic production data into geostatistical models. In this paper, we replace the gradient-based optimization in SSC by genetic algorithms (GA). GA, without requiring sensitivity, searches for global minimum. Although GA is computationally intensive, it provides significant flexibility to study parameters whose sensitivities are difficult to compute, e.g., master point locations. A steady-state GA is implemented under the SSC framework for searching the optimal master point locations, as well as the associated optimal perturbations that match the observed pressure, water cut and saturation data. We demonstrate that GA is easy to implement and results are robust. We examine different approaches of selecting master point locations including fixed, stratified random, and purely random methods. Results from this study demonstrate that there are not clear preferential master point locations that are best suited for matching production data for the given well pattern and for the given initial model. This is consistent with the early findings that master point locations can be randomly selected with the stratified random method yielding the best results due to its flexibility and good control for the overall model.

1 Introduction

Geostatistical reservoir models are widely used to model the heterogeneity of reservoir petrophysical properties, such as permeability and porosity. Geostatistical reservoir models must incorporate as much available, site-specific information as possible in order to reduce uncertainty in subsurface characterization, as well as in reservoir performance forecasting. Static data, such as core measurements, well logs, and seismic data, can relatively easily be integrated into geostatistical models using the traditional algorithms via conditional simulation (Deutsch and Journel, 1998). Integration of dynamic data, such as pressure, flow rate, fractional flow rate, and saturation data, is, however, a very difficult inverse problem and requires the solutions of the flow equations many times (Sun, 1994; Tarantola, 1987).

Geostatistically-based dynamic data integration has been an active area of research and a number of techniques have been reported in the literature (see Yeh, 1986 or Wen et. al., 1997 for review). The main objective is to match production data by modifying the initial geostatistical model in such a way that it preserves the underlying geostatistical

O. Leuangthong and C. V. Deutsch (eds.), Geostatistics Banff 2004, 691–701.

features built into the initial model, such as histogram, variogram, and other soft constraints.

The Sequential Self-Calibration (SSC) method has been shown to be very efficient and robust previously for integrating dynamic production data (Gomez-Hernandez et. al., 1997; Wen et. al., 1998, 2002). The SSC uses an optimization process to modify the original reservoir model. The efficiency of the SSC method comes from the master point concept and global updating. The master points are employed to reduce the number of parameters. Perturbations at the master points are then propagated into entire model to achieve the global updating.

Current implementation of SSC uses a gradient-based optimization to compute optimal perturbation values at the master points. This requires the calculation of sensitivity coefficients that measure the changes of reservoir flow responses with respect to the change of reservoir properties. In practice, the sensitivity calculations comprise the most CPU time in the inversion. A great deal of effort has been dedicated to speed up this calculation (e.g., Vasco et. al., 1998; Wen et. al., 1998, 2003). However, the sensitivities computed by these fast methods are often inaccurate which can cause difficulty in optimization. Also, gradient-based optimizations are often trapped by local minimums for highly nonlinear problems. Furthermore, there is a strong smoothing effect when successively adding perturbation field to the initial model at each iteration, resulting in a much smoother updated model than the initial one (Wen et. al., 2002).

The main goal of this work is to implement genetic algorithms (GA) for the optimization under the SSC framework. One advantage of using GA is that both master point locations and values of reservoir properties at the master points can be considered as variables in optimization. This allows us to compare the efficiency of different selecting schemes for master point locations, as well as to investigate the possibility of any preferential master point locations for the given problem.

2 Sequential Self-Calibration (SSC) Method

The SSC method was originally developed by Gomez-Hernandez and coworkers (Gomez-Hernandez et. al., 1997). The unique features of the SSC algorithm include (1) the concept of *master point* that reduces the parameter space to be estimated in optimization, (2) the propagation procedure through *kriging* that accounts for spatial correlation of perturbations, and (3) the fast computation of sensitivity coefficients within a single flow simulation run that makes inversion feasible. The main steps of the SSC method can be summarized as follows (Wen et. al., 1998, 2002):

a) Construct initial realizations: Multiple equal-probable initial property realizations are created by conventional geostatistical methods using specific histogram and variogram consistent with the data. If static (hard and soft) data are available, they should be honored with conditional simulation. Each realization is processed one at a time with the following steps.

b) Solve the flow equations for the current model using specific boundary and well conditions to obtain flow responses.

c) Compute the objective function that measures the mismatch between the observed production data and the flow solutions. If the objective function is smaller than a preselected tolerance, this realization is considered to honor the dynamic data and we move to the next realization. Otherwise, proceed to the following steps.

i) Select a few master locations (usually 1-3 per correlation range in each direction) and solve an optimization problem to find the optimal perturbations of reservoir property at these locations.

ii) Propagate the perturbations at master locations through the entire field by kriging the computed perturbations at master points. The model is then globally updated by adding the smooth kriged perturbation field to the previous model.

iii) Loop back to step b) until convergence or enough iterations have tried.

A gradient method was previously used in step i), which requires the sensitivity coefficients (derivatives) of flow responses with respect to reservoir property changes at the selected master locations. The method for computing sensitivity coefficients of pressure has been developed previously, i.e., they are computed as part of the flow simulation run (Gomez-Hernandez, et. al., 1997; Wen et. al., 1998). The sensitivity coefficients of water cut and saturation can be computed by a fast streamline-based approach, i.e., they can be obtained by simply book-keeping streamlines in the simulation field by using the 1D analytical solution along streamlines (Wen et. al., 2003).

In this paper, we apply genetic algorithms (GA) for optimization. The advantages of using GA include (1) no need to compute sensitivities, (2) global minimum, (3) easy to implement for different type of parameters, (4) easy to honor different type of constraints built in the initial model, and (5) CPU time does not significantly increase with the number of production data. When the gradient-based optimization is used, the outer iteration, step iii), was needed to account for the non-linearity between flow and parameters since we assume linear relation during the optimization process. By using GA, we do not need the outer iteration since there is not linear assumption. Thus only one global updating is needed and the smoothing effect by adding a smooth perturbation field is reduced to minimum.

3 Genetic Algorithms (GA)

Genetic algorithms (GA) belong to the group of artificial intelligence methods. Holland first introduced and applied the principles of evolution, such as genetic inheritance and Darwinian struggle for survival, for computation (Holland, 1975). He also showed that GA is remarkable in balancing exploration and exploitation of information to perform search. Since then, GA has been applied to many optimization problems (Goldberg, 1989).

To solve an optimization problem, GA manipulates a population of individuals that is randomly initialized. Each individual represents a potential solution to the problem. The quality of each individual is evaluated by a function or a process that assigns its “fitness” to the individual. Genetic operators are applied to a population to make it evolved toward a new and "better" one. This evolutionary process is repeated as many times as desired (number of generations). From a practical standpoint, genetic algorithms are assumed to provide, in the last generation, an enhanced population where some individuals-solutions ensure the convergence of the optimization problem. Applications of GA in inverse problems have been reported by Karpouzos et. al. (2001), Romero and Carter (2001), and Yu and Lee (2002).

We use a steady-state GA (DeJong, 1975) to search for the locations and the associated permeability values of a fixed number of master points that can minimize the mismatch between the flow simulation results and observed historical production data. Steady-state GA has an overlapping population where only a portion the population is replaced at each generation. The percentage that is replaced is specified by the GA users. The selection method is the traditional roulette wheel (fitness proportionate) selection. In this method, the probability of an individual to be chosen equals to the fitness of the individual divided by the sum of the finesses of all individuals in the population. Two genetic operators are used to generate offspring: uniform crossover and Gaussian mutation. Uniform crossover picks gene values from two parents randomly to compose the offspring. Figure 1 gives an example of two offspring that are created by uniform crossover. Gaussian mutation changes a gene value to a new value based on a Gaussian distribution around the original value.

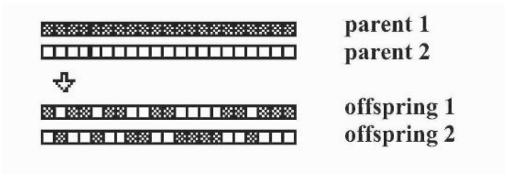

Figure 1. An example of uniform crossover.

4 Coupling SSC with GA

In this study, we only work on the permeability model. The objective function to be minimized is:

$$\boldsymbol{O} = \sum_{w_p=1}^{n_{wp}} \boldsymbol{W}_{w_p}\left[\hat{\boldsymbol{p}}(\boldsymbol{w}_p) - \boldsymbol{p}(\boldsymbol{w}_p)\right]^2 + \sum_{w_f=1}^{n_{wf}} \sum_{t_f=1}^{n_{tf}} \boldsymbol{W}_{w_f}\left[\hat{\boldsymbol{f}}(\boldsymbol{w}_f, \boldsymbol{t}_f) - \boldsymbol{f}(\boldsymbol{w}_f, \boldsymbol{t}_f)\right]^2 + \sum_{i=1}^{n_s} \boldsymbol{W}_{w_s}\left[\hat{\boldsymbol{s}}(\boldsymbol{i}) - \boldsymbol{s}(\boldsymbol{i})\right]^2 \quad (1)$$

where $\hat{p}(w_p)$ and $p(w_p)$ are the observed and simulated pressure at well w_p. $\hat{f}(w_f, t_f)$ and $f(w_f, t_f)$ are the observed and simulated water cuts at well w_f at time t_f. $\hat{s}(i)$ and $s(i)$ are the observed and simulated water saturation at cell i for the given time. W_{w_p}, W_{w_f} and W_{w_s} are the weights assigned to pressure, water cut, and water saturation to each well. n_{wp} and n_{wf} are the number of wells that have pressure and water cut data. n_{tf} is the number of time steps for water cut data. And n_s is the number of cells with water saturation data.

Following the SSC procedure as described above, for each initial reservoir model, we proceed the optimization process using GA. We first select a fixed number of master points and generate an initial population of initial master point locations and the associated permeability values. The master point locations are selected with three different methods: (1) fixed regular pattern, (2) stratified random (random within a regular coarse grid that covers the entire model), and (3) purely random (random within the entire model).

The permeability values at the master points are initially generated from a Gaussian function with the mean and variance consistent with the model. The constraints of the *ln(k)* values are the minimum and maximum limits. Note that if the *ln(k)* is not Gaussian, we can use a different distribution function. Also if we know the conditional pdf of each location, we can use such pdf to generate initial *ln(k)* values. This allows to honor different kind of constraints for the geostatistical model. Based on the generated *ln(k)* values at master point locations, as well as those at the initial model, we can compute the perturbations at the master point. We then interpolate the perturbation values at non-master point locations using kriging. An updated model is obtained by adding the perturbation field to the initial model. New fitness can be evaluated by solving flow using the updated model. If there are conditioning data that are already honored in the initial models, we include all conditioning data locations as master locations with zero perturbations. These master points are included in the GA searching process. Instead, they are simply added at the interpolation step.

With *Nm* master points (excluding the conditioning data points), the GA genome is an array of *Nm* integer and *Nm* real numbers. They are the locations and *ln(k)* values for the *Nm* master points. The steady-state GA searches for the new master point locations and the associated permeability values until the mismatch between the flow simulation results and observed historical production data is minimized. We retain the best individual (the optimal master point locations and the associated *ln(k)* values) at the end of the GA.

5 An Example

In this section, we demonstrate the applications of the coupled SSC/GA method for constructing reservoir permeability models from pressure, water cut and water saturation data using a synthetic data set. In the example, we assume porosity is known and constant as $\phi = 0.2$.

Figure 2(a) shows a 2-D geostatistical reference field (50x50 grid with cell size 80 feet x 80 feet). The model is generated using the Sequential Gaussian Simulation method (Deutsch and Journel, 1998). The *ln(k)* has Gaussian histogram with mean and variance of 6.0 and 3.0, respectively. The unit of permeability (*k*) is milli Darcy. The variogram is spherical with range of 800 feet and 160 feet in the direction of 45 degree and 135 degree, respectively. We assume an injection well (I) at the center of the model with 4 production wells (P1 to P4) at the 4 corners. The injection rate at the injection well (I) is 1600 STB/day and the production rate for the 4 production wells is 400

STB/day/well. The thickness of the reservoir is assumed constant of 100 feet. All four boundaries are no-flow boundaries. The initial pressure is constant at 3000 psi for the entire field.

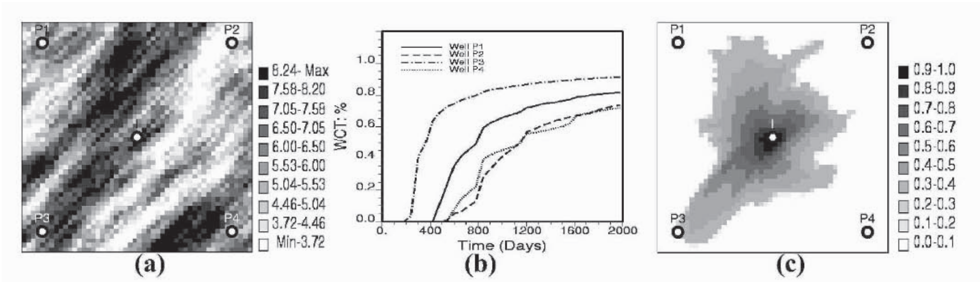

Figure 2. (a) The 2-D reference log-permeability field, (b) water cuts from the 4 production wells, and (c) water saturation distribution at 400 days.

The main features of this reference field are: (1) a high permeability zone and a low permeability zone in the middle of the field, (2) high interconnectivity between well I and well P3, (3) low interconnectivity between well I and wells P2 and P4. This reference field is considered as the true model, and our goal is to reconstruct reservoir models based on some production data that are as close to this true field as possible.

The reservoir is initially saturated with oil. Water injection and production are solved using a streamline simulator for 2000 days. Mobility ratio is 10 and standard quadratic relative permeability curves are used with zero residual saturation for oil and water. Compressibility and capillary pressure are ignored. Pressure field is updated every 400 days to account for the change of mobility during the streamline simulation. We assume the observed production data are: (1) bottom hole pressure (BHP) of each well at the end of the simulation, (2) water cut history of each production well, and (3) water saturation distribution of entire model at 400 days. These production data are supposed to mimic the practical situation of a producing field with 4D seismic survey. The "observed" BHP for I and P1-P4 are given in Table 1, the water cuts of 4 production wells and water saturation distribution at 400 days are given in Figures 1(b) and 1(c), respectively. Note that the fast water breakthrough at well P3 and late breakthrough at wells P2 and P4.

We generate multiple initial realizations using the same histogram and variogram as the reference field. These initial models are then modified to match the observed production data using the coupled SSC/GA method. We use 25 master points that are selected stratified randomly within each of the 5x5 coarse grid cells (each coarse cell represents a 10x10 fine cells).

The population size in GA is 50 and the maximum number of generations evolved is 50. The crossover rate is 90% while the mutation rate is 1%. This means that the selected two parents have 90% to be cross-over with each other to produce 2 offspring. The produced offspring (regardless if crossover has been performed or not) have 1% to be mutated. In other words, two offspring can be the results of crossover and mutation, crossover only, mutation only or identical copies. Among the population of 50 individuals, the worst 60% will be replaced with the new offspring. This is the steady-state GA explained in Section 3.

All 50 models in the last generation closely match the production data (see Figure 5). The best individual at the last generation is chosen as the final updated model. Figure 3 shows two initial permeability fields (top row) and the resulting master point locations (plus) and the perturbation fields (middle row). The final updated models are shown at the bottom of the figure. The BHPs at wells computed from the initial and updated models are given in Table 1. The water cut and water saturation matches from the initial and updated models are given in Figures 4 and 5.

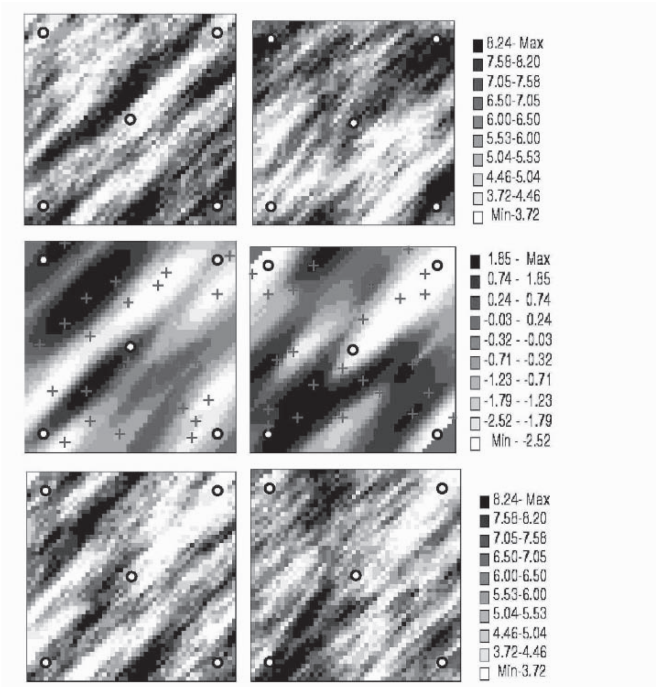

Figure 3. Two initial realizations of *ln(k)* model (top), the computed perturbation fields and master point locations (middle), and the resulting updated models (bottom).

Compared to the reference field, we can see that the spatial variation patterns in the two initial models are quite different from the reference model, resulting in significant deviations of flow responses from the "observed" production data. After inversion, the updated models display spatial variation features very similar to the reference model with flow results matching the "observed" data closely (see Figures 4 and 5). Particularly, in both models, in order to match the production data, permeabilities in the region between wells I and P3 are increased, while in the region between wells I and P2, permeabilities are reduced (see Figure 3). Based on these, we can conclude that the GA is capable of finding the optimal master point locations, as well as the associated optimal permeability values that match the production data.

Well	I	P1	P2	P3	P4
Reference	3043	2985	2468	3022	2917
Initial, #1	3135	2489	2920	2995	2974
Updated, #1	3071	2950	2422	3016	2918
Initial, #2	3037	3007	3022	2925	2872
Updated, #2	3055	2949	2505	3019	2929

Table 1. Comparison of BHPs from the two initial and updated models with the reference field

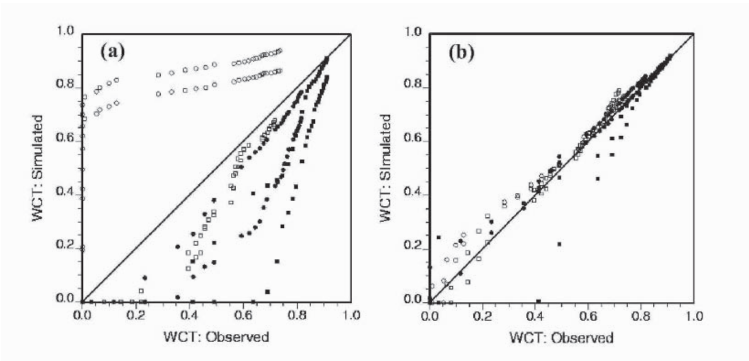

Figure 4. Scatter plots of water cuts from the two initial and updated models with respect to the observed data: (a) initial models; (b) updated models. Open circles: W1, filled circles: W2, open squares: W3, filled squares: W4.

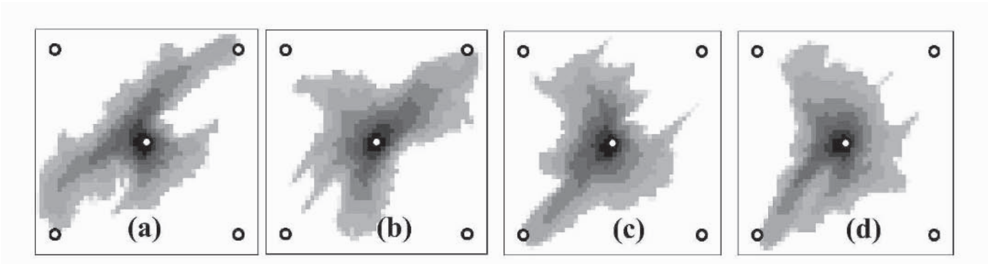

Figure 5. Water saturation distributions from the two initial and updated models: (a) initial model 1; (b) initial model 2; (c) updated model 1; (d) updated model 2. Note the reference water saturation distribution is in Figure 2(c).

Figure 6 shows the changes of objective function at each generation during the GA operation for the first model indicating the rapid reduction of objective function. The total number of function evaluation (flow simulation run) for generating one realization is about 1450.

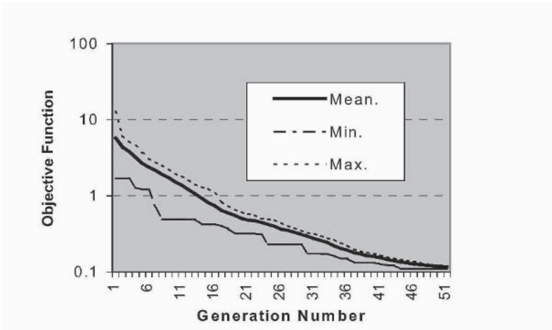

Figure 6. Variation of objective function with generation in GA.

Similar results are obtained (not shown here) by using fixed or purely random master point locations. Using fixed or pure random master point locations, however, yields the updated models with slightly larger objective function than using the stratified random method. This can be explained by (a) fixed master point locations do not provide enough flexibility on selecting best locations, (b) purely random master point locations may not provide enough overall coverage of the entire model, and (c) the stratified random method provide best compromise between the overall coverage and flexibility.

6 Discussion

Traditional pilot point method seeks the “best” pilot point locations based on sensitivity coefficients and then computes the optimal perturbations at these locations (RamaRao et. al., 1995). New “best” pilot point locations are added after each iteration. A significantly amount of CPU time is required for searching the “best” pilot point locations. The SSC method, however, uses a fixed number of randomly selected master points and computes the optimal perturbations at those locations (Wen et. al., 1998, 2002). Master point locations are updated after each several iterations during the inversion. This eliminates the time-consuming step of searching “best” locations as in the traditional pilot point method. Using the coupled SSC/GA method, one interesting issue is to investigate that, for a given well pattern or a given initial model, if there exist preferential master point locations that are superior to other locations for matching the given production data.

Figure 7 presents the total number of times that a particular cell is selected as master point from the 100 realizations using the stratified or purely random method. Clearly, there is no spatial pattern that is noticeable from these maps. Instead, they look like more or less random noise in the entire model without any structure. This demonstrates that, from a statistical point of view, there is no preferential locations that are better suited for being master point locations for the given well configurations, as long as the master points are not overly clustered in the space.

To investigate the possibility of any preferential master point locations for a given initial model, we update the same initial model 100 times using different random number seeds resulting in 100 updated models and 100 sets of master point locations. Using the two initial models as shown in Figure 3, Figure 8 shows the total number of times that a particular cell is selected as master point from the 100 runs using the stratified and purely random methods. In this case, we can see that there is slightly higher tendency that the master points are selected at areas where initial values are either too high or too low. This displays the efficiency of our method to pick up the right places to update the model. Nevertheless, this tendency is not significant indicating that there are no specific locations that are significantly better as master locations for a given initial model. Our results (not shown here) also indicate that the stratified random method provides best results in terms of the accuracy in data matching, resulting in updated reservoir models with less uncertainty compared to the fixed or purely random master point locations. From above investigation, we can conclude that master (pilot) point locations are not critical for the SSC inversion. There are no such locations that are "best" as master/pilot point locations for a given problem. Master points can be selected randomly provided that they can cover the overall model space for the given correlation structure.

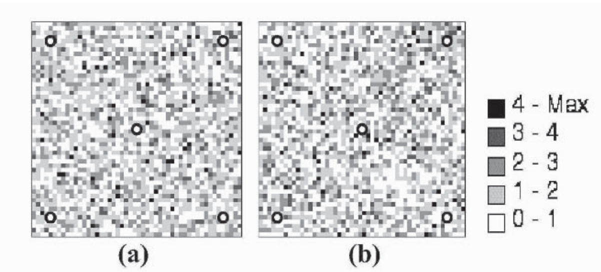

Figure 7. Total number of times that a cell is selected as master point for 100 realizations: (a) stratified random method, (b) purely random method.

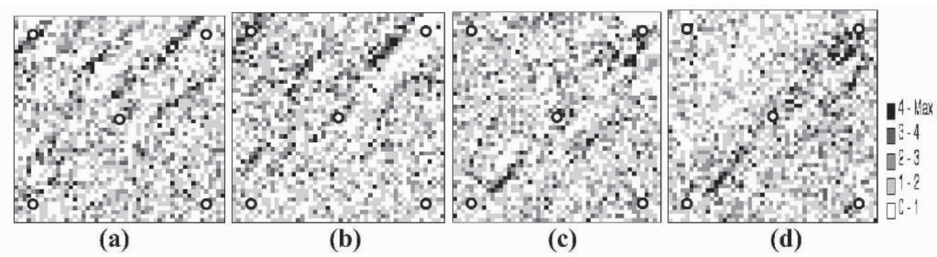

Figure 8. Total number of times that a cell is selected as master point from 100 runs using the two initial models: (a) initial model 1 and stratified random method; (b) initial model 1 and purely random method; (c) initial model 2 and stratified random method; (d)) initial model 2 and purely random method

7 Summary and Conclusions

We implemented a steady state GA under the SSC framework as an optimization process replacing the original gradient based optimization procedure. The coupled SSC/GA method is used to invert geostatistical reservoir permeability model from dynamic production data. The results show that the coupled SSC/GA is capable of finding optimal master point locations and the associated optimal perturbations within a reasonable number of generations. The results are accurate and robust.

GA allows us to investigate whether or not the “best” master point locations exist for a particular well pattern and for a particular initial model. We showed that there is no clear tendency with respect to where the master points should be, i.e., there are not preferential locations where the “best” master locations can be chosen. In other words, inversion results are not sensitive to what locations are chosen as master points. Master point locations can be selected randomly as long as they cover the entire model. This provides explanation to the previous studies on why randomly selected master points yielded similar results to those using “carefully” selected master points.

References

DeJong, K. A., *An Analysis of The Behavior of a Class of Genetic Adaptive Systems*, Ph.D. Thesis, University of Michigan, 1975.

Deutsch, C. V. and Journel, A. G., *GSLIB: Geostatistical Software Library and User’s Guide*, 2nd edition, Oxford University Press, 1998.

Goldberg, David E. *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Pub. Co., 1989.

Gomez-Hernandez, J. J., Sahuquillo, A. and Capilla, J. E., Stochastic Simulation of Transmissivity Fields Conditional to Both Transmissivity and Piezometric Data, 1. The theory, *Journal of Hydrology*, 203(1-4) 1997, p. 162-174.

Holland, J., *Adaptation in Natural and Artificial System*, Univ. of Mich. Press, 1975.

Karpouzos, D. K., Delay, F., Katsifarakis, K. L. and De Marsily, G., A Multipopulation Genetic Algorithm to Solve the Inverse Problem in Hydrogeology, *Water Resour. Res.*, Vol. 37 , no. 9, 2001, p. 2291-2302.

RamaRao, B.S., LaVenue, A.M., de Marsily, G. and. Marietta, M.G, Pilot Point Methodology for Automated Calibration of an Ensemble of Conditionally Simulated Transmissivity Fields, 1. Theory and Computational Experiments, *Water Resour. Res.*, 31, no. 3, 1995, p. 475-493.

Romero, C. E. and Carter, J. N., Using Genetic Algorithms for Reservoir Characterization, *Journal of Petroleum Science and engineering*, vol. 31, 2001, p. 113.

Sun N.-Z., *Inverse Problem In Groundwater Modeling*, Kluwer Academic Publishers, 1994.

Tarantola, H., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, 1987.

Vasco, D. W., Yoon, S., and A. Datta-Gupta, Integrating Dynamic Data Into High-Resolution Reservoir Models Using Streamline-Based Analytical Sensitivity Coefficients*, SPE paper 49002 presented at the 1998 SPE Annual Technical Conference and Exhibition*, New Orleans, LA, Sept. 27-30, 1998.

Wen, X.-H, Deutsch, C. V. and. Cullick, A.S, A Review of Current Approaches to Integrate Flow Production Data in Geological Modeling, *In Report 10, Stanford Center for Reservoir Forecasting*, Stanford, California, 1997.

Wen, X.-H., Deutsch, C. V. and Cullick, A.S, High Resolution Reservoir Models Integrating Multiple-Well Production Data, *SPE Journal*, December, 1998, p. 344-355.

Wen, X. H., Deutsch, C. V. and Cullick, A. S., Construction of Geostatistical Aquifer Models Integrating Dynamic Flow and Tracer Data Using Inverse Technique, *J. of Hydrology*, vol. 255, 2002, p.151-168.

Wen, X. H., Deutsch, C. V. and Cullick, A. S., Inversion of Dynamic Production Data for Permeability: Fast Streamline-Based Computation of Sensitivity Coefficients of Fractional Flow Rate, *J. of Hydrology,* vol. 281, 2003, p. 296-312.

Yeh, W. W.-G., 1986. Review of parameter identification procedure in groundwater hydrology: The inverse problem. *Water Resour. Res.*, vol. 22, no. 2, 1986, p. 85-92.

Yu, T and Lee, S., Evolving Cellular Automata to Model Fluid Flow in Porous Media*, Proceedings of the 2002 NASA/DoD Conference on Evolvable Hardware*, IEEE Press, 2002, p. 210-218.