2006 IEEE Congress on Evolutionary Computation
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

A Hybrid of Sequential-Self Calibration and Genetic Algorithm Inversion Technique for Geostatistical Reservoir Modeling

Tina Yu, Xian-Huan Wen and Seong Lee

Abstract— **Geostatistical modeling is a widely used approach to model the heterogeneity of reservoir petrophysical properties. This paper investigates a geostatistical-based inversion technique, a hybrid of sequential-self calibration (SSC) and genetic algorithm (GA), to model reservoir permeability. In this method, a GA is used to search the optimal master point locations, as well as the associated optimal permeability. These permeability values are then propagated to the entire reservoir using Kriging algorithm to match the dynamic production data.**

We demonstrate that GA is easy to implement and the results are robust. Additionally, we experimented with various numbers of master points, including a linked-list genotype which permits a flexible number of master points. The results show that GA is able to find various numbers of master points and their locations that are suitable for the reservoir field we studied. These numbers are within a small range and are sufficient to capture the heterogeneity of the reservoir permeability to match the production data.

The ability of the SSC-GA method to model reservoir permeability by simultaneously optimizing *the number of master points*, *the locations of these master points* and *the associated permeability* in this case study suggests that the technique might be effective with other larger fields.

I. Introduction

Geostatistical modeling refers to the process of generating earth subsurface models using statistical methods. In petroleum industry, geostatistical modeling is a widely used approach to model the heterogeneity of reservoir petrophysical properties, such as permeability and porosity. This approach incorporates both static and dynamic data to build reservoir models. Static data, such as core measurements, well logs and seismic data, are relatively easy to integrate using traditional statistical algorithms and conditional simulation [4]. In contrast, the integration of dynamic data, such as pressure, flow rate and saturation data, requires multiple computer flow simulations. The integration of dynamic data to reservoir models is a inverse problem, which is known to be difficulty to solve [8], [9].

To address this inverse problem, various geostatistical-based inversion techniques have been developed [13]. The main concept behind these techniques is to update the initial geostatistical model to match dynamic data, *without disrupting* the underlying geostatistical features built into the initial model, such as histogram, variogram, and other soft constraints. This differs from the traditional reservoir history matching process where *local or regional multipliers* to reservoir properties are adjusted to match dynamic data *without the consideration* of the underlying geostatistical features in the original model. Such approach can potentially cause the discontinuities inside the reservoir and may destroy the correlation between geological features built into the initial reservoir models. Since geostatistical-based inversion techniques do not have these shortcomings, it is argued that they can generate more realistic reservoir models.

Previously, we have investigated a geostatistical-based inversion technique: a hybrid of sequential-self calibration (SSC) and genetic algorithm (GA) method [11], [12]. In this technique, a few number of master point locations in the reservoir field are selected and their associated reservoir property values identified. These property values are then propagated to the entire reservoir using the statistical Kriging algorithm. In order to construct reservoir models that match the dynamic data, two critical questions need to be addressed. First, what are the locations of these master points? Second, what are their associated reservoir property values?

We have applied GA to search for the answers to both questions. Indeed, as a global search algorithm, GA is able to find both sets of parameters (master point locations and their associated property values) to update the reservoir models to the dynamic data. In [11], we reported that the locations of the master points are not important in the well patterns we studied. The SSC-GA technique can find good models to match the production data using any randomly selected locations, as long as the locations are not overly clustered in one particular area of the field. In another work [12], we showed that the SSC-GA technique is capable of identifying important large-scale spatial variation patterns (e.g. well connectivity, near well averages, high flow channels and low flow barriers) embedded in the reservoir heterogeneity.

This paper investigates the number of master points in the SSC-GA method. In particular, we address the question "Does the number of master points in the reservoir model have impact on the SSC-GA modeling results?"

We organize the rest of the paper as follows. Section II presents the hybrid SSC-GA technique. In Section III, the synthetic oil field used in this study is described. Section IV gives the setup to run GA experiments. The results are then reported in Section V. We discuss our findings and address the posted question in Section VI. Finally, Section VII concludes the paper.

Tina Yu is with the Department of Computer Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, CANADA (phone: 709-737-6943; fax: 709-737-2009; email: tinayu@cs.mun.ca).

Xian-Huan Wen and Seong Lee are with Chevron Energy Technology Company, San Ramon, CA 94583 USA (email: wenxh@cnooc.com.cn and seon@chevron.com).

II. The Hybrid SSC-GA Inversion Technique

Sequential self-calibration (SSC) method is a geostatistical-based inverse technique that allows for fast integration of dynamic production data into geostatistical models. The method was originally developed by Gomez-Hernandez and coworkers [5] and later expanded by others [6]. The main concept of SSC algorithm is intelligent selection of master point locations and the propagation of reservoir properties from these master point locations to the entire field. In this way, there is no need to model reservoir properties at the cell level, which is very time consuming. Instead, only the master point locations and the optimal perturbations for propagation to match the dynamic production data need to be identified.

This concept has been tested by incorporating a gradient method [10] and a genetic algorithm (GA) [11] to select master point locations and to optimize perturbations. The results show that GA outperforms gradient method and enables the SSC technique to incorporate dynamic data, such as fluid flow and pressures, to geostatistical models more effectively [12].

A. Work Flow

Figure 1 gives the work flow of the hybrid SSC-GA technique. Initially, multiple equally-probable initial reservoir models are created. The reservoir properties in these initial models are generated by conventional geostatistical methods using the specific histogram and variogram that are consistent with the data. If static (hard and soft) data are available, they should be honored with conditional simulation.

For each of the *N* initial reservoir models, a GA is applied to update the model to match the dynamic data:

- Create a population of individuals, where each individual is a vector consists of possible master point locations and their associated property values. The property values are generated based on a Gaussian function with the mean and variance that are the same as that used to create the initial model.
- Evaluate the fitness of each individual model in the population with the following steps:
  - Compute the perturbations at the master points, based on their reservoir property values and the values in the initial model;
  - Interpolate the perturbations at the non-master point locations using Kriging algorithm;
  - Obtain an updated reservoir model by adding the perturbations to the initial model.
  - Conduct computer simulation on the updated model by solving the flow equations using specific boundary and well conditions to obtain flow responses;
  - Calculate the mismatch between the flow responses and the dynamic production data. This mismatch is the fitness of the model. The smaller the mismatch is, the fitter the reservoir model is.
  - Apply selection, crossover and mutation to generate a new population of individual models.

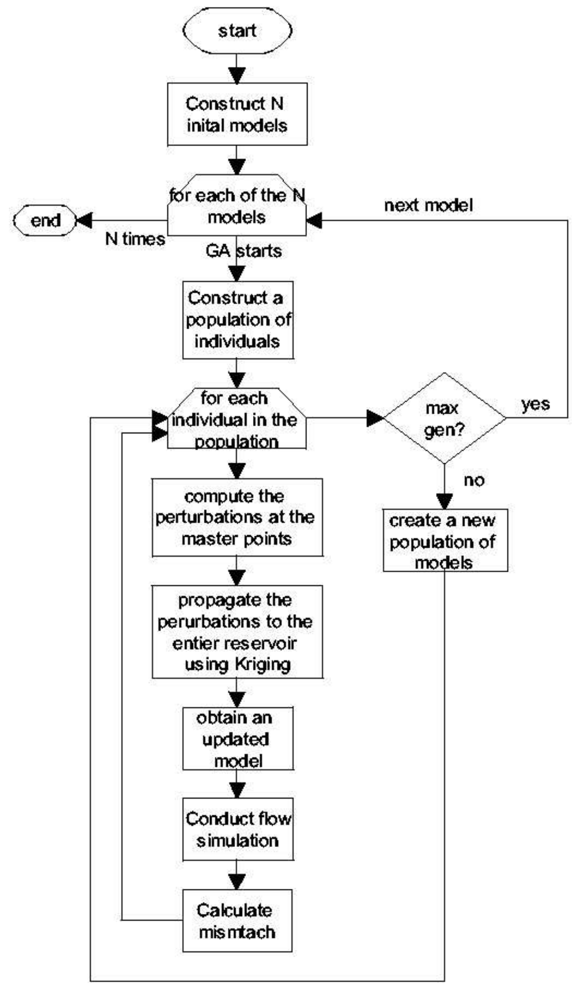

Fig. 1. The hybrid SSC-GA technique work fbw.

  - Loop back to the *compute the perturbations* step until either the population converges or the maximum number of generation is reached.

For *N* number of initial models, *N* number of GA runs are conducted, each of which delivers a population of updated models at the end of the GA run. Among them, the reservoir models with the smallest mismatch between simulation flow responses and the dynamic production data is chosen is the final model.

III. Case Study

We used an artificial reservoir field generated from computer simulation to study the number of master points in the SSC-GA inverse technique. In this way, all data (statistic and dynamic) are genuine without noise. The reservoir property modeled by the SSC-GA is permeability.

In this field, there are 5 wells. On injection well (*I*) is at the center while 4 other production wells (*P1* to *P4*) are at the 4 corners of the field. 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. Additionally, we made the following assumptions:

- The thickness of the reservoir is 100 feet across the entire field.
- All four boundaries of the field are no-flow boundaries.
- The initial pressure is constant at 3000 psi for the entire field.
- The porosity ϕ of the reservoir is 0.2 across the entire field.

With the above assumptions, we applied the Sequential Gaussian Simulation (SGS) method [3] to generate the reference reservoir model. Figure 2 gives the simulated 2-D model in 50x50 grids, where each cell is of size 80 feet x 80 feet. The permeability of each cell is represented by different gray scaled colors. The permeability (k) unit is milliDarcy. The mean and variance of the Gaussian histogram of *ln(k)* are 6.0 and 3.0, respectively. The variogram is spherical with range of 800 feet and 160 feet in the direction of 45 degree and 135 degree, respectively.

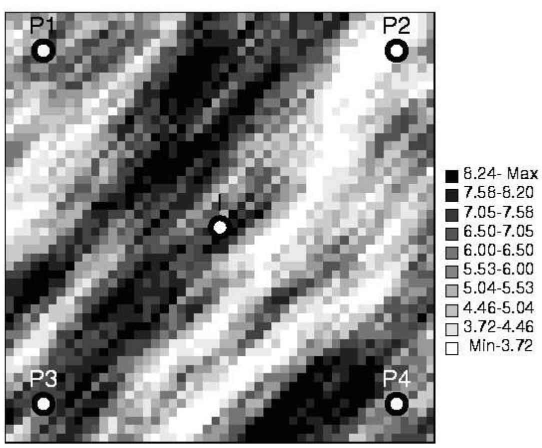

Fig. 2. The 2-D reference model, where permeability of each cell is generated using the Sequential Gaussian Simulation Method.

The main features of this reference field are:

- There is a high permeability zone and a low permeability zone in the middle of the field;
- Interconnectivity between the injection well *I* and the production well *P3* is high;
- Interconnectivity between the injection well *I* and the production wells *P2* and *P4* is low.

This reference field is considered as the true model. We will apply the SSC-GA technique to reconstruct this reservoir model by incorporating three sets of dynamic production data.

A. Dynamic Production Data

The dynamic production data are generated from this reference model through computer simulation. The data include:

- water cut (WC) history of each production well;
- water saturation (WS) distribution of the entire reservoir at the end of the last 400 days;
- bottom-hole pressure (BHP) of each well at the end of the simulation.

We used the following procedures to generate these 3 sets of data. Initially, the reservoir was saturated with oil. Water injection and production are generated using a streamline simulator for 2000 days. The mobility ratio used is 10 and the standard quadratic relative permeability curves used are with zero residual saturation for oil and water. The pressure field is updated every 400 days to account for the change of mobility during the streamline simulation. Compressibility and capillary pressure are ignored in the simulation.

The generated water cuts for each of the 4 production wells are shown in Figure 3. Note that the production well *P3* has fast water breakthrough while production wells *P2* and *P4* have late water breakthrough. The generated water saturation distribution at the end of the last 400 days are given in Figures 4. The generated BHP for injection well *I* and the production wells *P1* to *P4* are given in Table I.

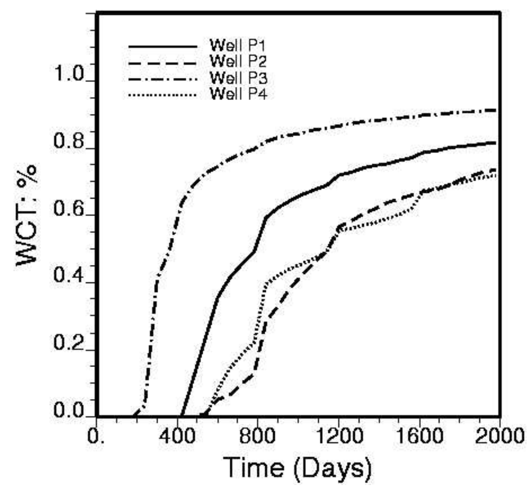

Fig. 3. Water cuts data of the reference model.

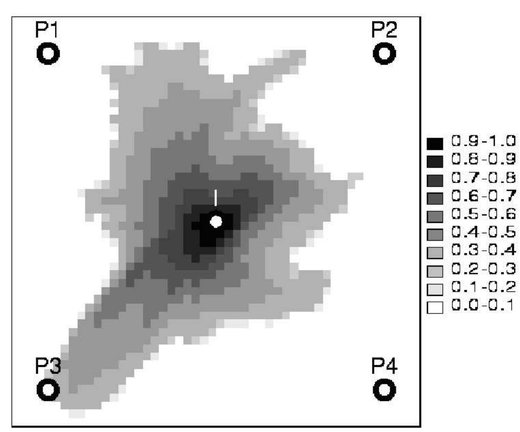

Fig. 4. Water saturation distribution of the reference model.

TABLE I

BOTTOM HOLE PRESSURE FOR THE 5 WELLS.

Well	I	P1	P2	P3	P4
BHP	3034	2985	2468	3022	2917

IV. EXPERIMENTAL SETUP

We first created 100 initial models based on the same histogram and variogram as that used to generate the reference field. These initial models are then updated by the hybrid SSC/GA method described in Section II to match the observed 3 sets of production data.

A. Genotype

The GA genotype is a mix-typed vector containing the locations of the master points and their associated permeability values. The data type of master point locations is integer while the data type for permeability is real number. Each genotype has length 2*N*, where *N* is the number of master points. In this study, we experimented with 3 different *N* values: 15, 25, 50. Additionally, we conducted one experiment using a flexible-length genotype (a linked-list data structure). The value of *N* is between 1 and 2,500 and is decided by GA.

The linked-list is implemented as an one-dimension array of 5,000 elements, indexed by integers from 0 to 4,999. The contents indexed by an odd number are permeability while the contents indexed by an even number are links to the next master point location. Figure 5 gives an example of the linked-list genotype.

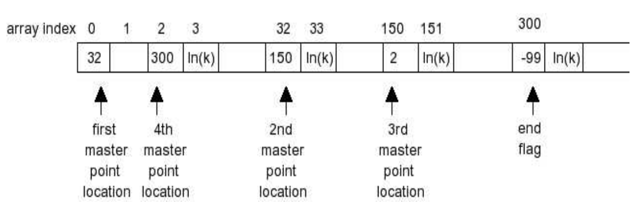

Fig. 5. An example of the linked-list genotype.

The first element in the array (index 0), is the first master point location (32). The associated permeability is indexed by 33. The next master point location is 150 (indexed by 32). Its associated permeability is indexed by 151. The rest of the master point locations and their associated permeability can be traced using the same procedure. The linked-list ends when the end-flag (-99) is encountered.

Except the end-flag (-99), all elements in the linked-list are subject to GA crossover and mutation operations. When any of the the indices to the next master point locations is modified, the data structure changes and the number of master points also changes. In other words, GA simultaneously optimizes 3 different variables to update the reservoir model: *the number of master points*, *the locations of these master points* and *associated permeability at these master point locations*.

B. Initialization

Regardless of the size of *N*, the master points locations in the initial population are selected among the 2,500 possible cell locations using the stratified random method. The initialization of permeability values at each master point location is based on a Gaussian function with the same mean and variance as that used to generate the initial model. As a convention, permeability *k* is represented in log scale (*ln(k)*). The *ln(k)* values are constrained to be between 0.804 and 11.196 throughout the GA runs. Note that if the *ln(k)* is not Gaussian, we can use a different distribution function. Also if we know the conditional Probability Density Function (PDF) at any of the locations, we can use that PDF to generate the initial *ln(k)* value at that location. In this way, different kind of constraints can be honored in the initial geostatistical models.

C. Fitness Function

The fitness function is the mismatch between the production data (also called observed data) as described in Section III and the simulation results from the reservoir model updated using the genotype information (see the work flow in Section II). Equation 1 gives the fitness function:

$$
\begin{aligned}
F = & \sum_{i=1}^{m}\sum_{j=1}^{o} W_f[\hat{f}(w_i,t_j) - f(w_i,t_j)]^2 \\
& + \sum_{i=1}^{n} W_p[\hat{p}(w_i) - p(w_i)]^2 \\
& + \sum_{i=1}^{q} W_s[\hat{s}(i) - s(i)]^2
\end{aligned}
\tag{1}
$$

where $\hat{f}(w_i,t_j)$ and $f(w_i,t_j)$ are the observed and simulated *water cuts* at well i at time j. The observed and simulated *pressure* at well i are $\hat{p}(w_i)$ and $p(w_i)$, respectively. The observed and simulated *water saturation* at cell i for a given time are $\hat{s}(i)$ and $s(i)$, respectively. Each of the mismatch of water cut, pressure and water saturation are assigned with a different weight, represented as W_f, W_p and W_s. In this study, W_f is 1, W_p is 10 and W_s is 1. There are m wells with water cut data and n wells with pressure data. The number of time steps for water cut data is o and the number of cells with water saturation data is q.

D. Genetic Operators and Parameters

The GA is steady-state [1] with 60% replacement rate. This means that at each generation, the better 40% of the population remains while the other 60% is replaced by newly created offspring. New offspring always make into the population, regardless of whether their fitness are better than the worse 60% of the original population or not.

We used the traditional roulette wheel (fitness proportionate) method to select winners for reproduction. In this method, the probability of an individual to be chosen is the fitness of the individual divided by the sum of the fitness 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 both parents randomly to compose the offspring. Gaussian mutation changes a gene value to a new value based on a Gaussian distribution around the original value.

We do not know whether this is the best combination of genetic operators. In general, uniform crossover is more disruptive than one-point and two-point crossovers. However, for small population size, which does not provide the necessary sampling accuracy, this disruption gives the exploration needed for adaptive search [2]. In this problem, the size of the search space is unbounded (the permeability values are real numbers). To get enough sampling accuracy, it needs a large population size, which is not possible because flow simulation is very time-consuming. We therefore choose uniform crossover and Gaussian mutation to work with a small population size (50). Table II gives the GA parameter values used to run the experiments.

TABLE II

GA PARAMETER VALUES.

Parameter	Value
population size	50
number of generation	55
crossover rate	90%
mutation rate	1%
number of runs	100

V. RESULTS

When a different number of master points is used, the SSC-GA method delivered a different result. However, regardless of the number of master points used, all 50 models in the GA last generation give a similar match to the production data. Figure 6 shows the typical progress of a GA run. Before generation 30, the fitness of the population improved quickly. After that, only small improvements were observed. At the end of the run, the population average fitness is very similar to the best model fitness, which is also very similar to the worst model fitness. This indicates that the population has converged with all genotypes having similar contents. We therefore chose the best model at the last generation as the final model.

Figure 7 gives the fitness of the final models from all 100 runs, under different master point implementations. It is clear that 15 master points are not sufficient to capture the heterogeneity of the reservoir permeability, hence give poor match to the production data. When the number of master points is increased to 25, the performance improved. The best results come from 50 master points implementation.

In Table III, we give the average fitness of the 100 runs for the different number of master points implementations. We have also computed their upper and lower bounds with 99% confident interval, which are shown graphically in Figure 8.

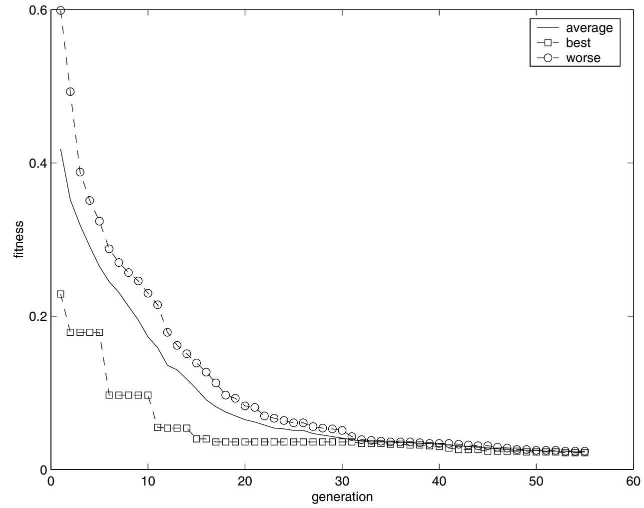

Fig. 6. The typical progress of a GA run.

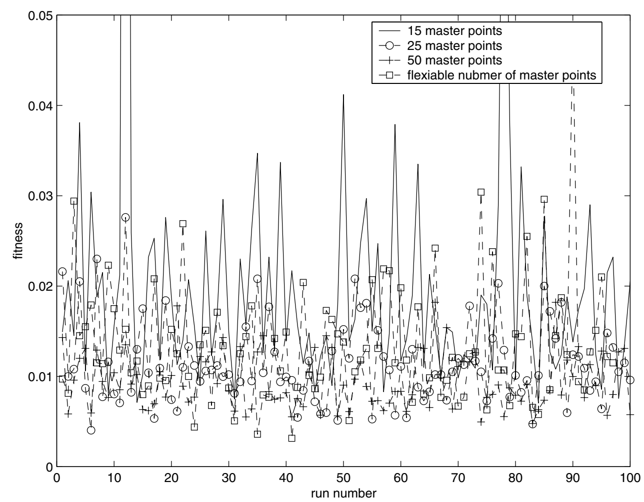

Fig. 7. Fitness of the final models using different number of master points.

It is evident that a larger number of master points gives a finer scale of permeability modeling, hence delivers better results. However, they also increase the dimension of the genotype and the size of search space. When the genotype is a linked-list structure, which provides up to 2,500 master points, SSC-GA did not deliver better results than the results of 50 master points implementation (more discussion in Section VI). Of course we should not forget that the main concept of SSC is to use a small number of master points to model reservoir properties. If we refine the the modeling scale to the cell level, this is no longer a SSC technique.

VI. ANALYSIS AND DISCUSSIONS

Although the linked-list genotype allows the SSC-GA to select a maximum of 2,500 master points, the 100 final models contain only 10 to 30 master points. As mentioned in Section V, increase the number of master points also increase the size of the search space and make the optimization task harder. For the permeability pattern in this field, 30

TABLE III
SSC-GA EXPERIMENTAL RESULTS.

no. of masters	mean	stdev	t-table	upper	lower
15	0.0206	0.021	2.6264	0.02	0.0211
25	0.0114	0.00456	2.6264	0.0113	0.0115
50	0.00949	0.00312	2.6264	0.0094	0.00957
*	0.0136	0.00685	2.6264	0.0134	0.0137

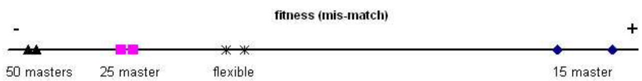

Fig. 8. Average fitness of 100 runs and their intervals with 99% confidence.

master points are sufficient to capture the heterogeneity of the field. The experimental results show that the SSC-GA technique was able to select a sufficient number of master points to identify such pattern. This indicates that SSC-GA is capable of simultaneously optimizing the number of master points, the locations of these master points and the associated permeability at these master point locations to match the production data successfully.

Although the overall results from the linked-list genotype representation are not as good as the results from the 50 master points implementation, many of the final models from the linked-list representation are better than the best model from the 50 master points implementation. Some examples are run 41, run 35 and run 24 (see Figure 7).

This suggests that by hard-wiring the number of master points, SSC-GA loses some exploration power to find better models with a different number of master points. However, such constraint also ensures that the search is within a defined area of the search space. When the defined area is good, which is the case when 50 master points were used, the results are good and the average results have a tight confidence interval (see Figure 8). This trade-off between exploitation and exploration is a common dilemma in all optimization techniques. For optimization problems where the results require high certainty, exploitation by using constraints can deliver better results. On the contrary, if finding the best model is the goal, exploration of flexible representation may satisfy the objective better.

Are good models associated with a certain number of master points? In other words, is the fitness of the model correlated with the number of master points? Figure 9 shows that such correlation does not exist. For example, both models from run 93 and run 90 have 27 master points. Yet, their fitness values are far apart. This indicates that there is no single best number of master points to model the permeability of the studied field. As long as the number is large enough to capture the heterogeneity of the reservoir property, a range of different number of master points can deliver similar good results.

Figure 10 gives 2 final models that have different numbers of master points, one has 24 and the other has 26. However, the permeability at these master points are similar to the reference filed (Figure 2). As a result, the globally updated model matches the reference field very closely.

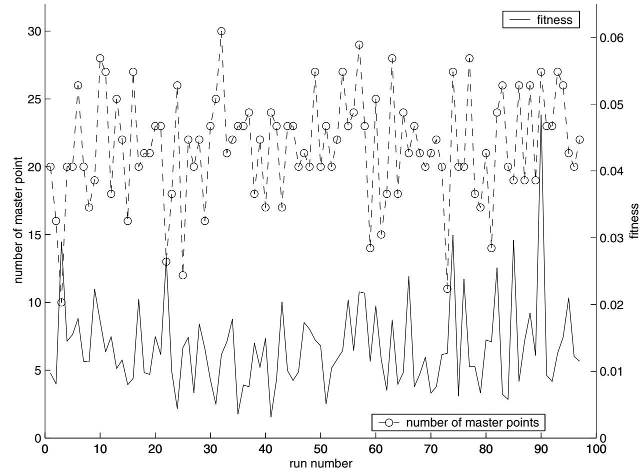

Fig. 9. Fitness vs. number of master points using a flexible length genotype representation.

In [11], we reported that the locations of the master points are not important for SSC-GA to model the permeability for the same reservoir field. The SSC-GA technique can find good model to match production data using any randomly selected master point locations, as long as the locations are not overly clustered in one particular ares of the field. This work shows a similar result: there is no one best number of master points to model the reservoir permeability. The GA with a linked-list genotype representation is able to find a suitable number of master points and delivers good models that match the reference field.

The computation time used for the GA runs is not most economical. Each set of 100 runs took about 8 hours to complete, using a single Pentium CPU machine. However, as the computer power increases so rapidly, we anticipate a speedy decrease of the modeling time. Nevertheless, in order to apply the technique to real reservoir fields, a more efficient fitness evaluation scheme is needed. Currently, developing surrogates for computational expensive fitness function to accelerate evolutionary optimization is an active research area [14], [7]. We are optimistic about applying the SSC-GA technique to the real world reservoir field in the near future.

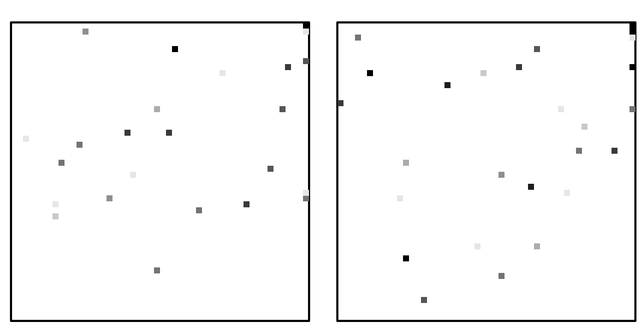

Fig. 10. Master point locations and their associated permeability for two final models with a different number of master points.

VII. Concluding Remarks

The concept of intelligent master points selection and then applying statistical algorithm to propagate properties from those master point locations is a very appealing approach for reservoir modeling. This is because experimental studies indicate that the properties of earth surface follow a certain kind of statistical distributions [3]. Additionally, reservoir modeling in fine scale is still computationally too expensive. This alternative approach can provide useful reservoir information within a reasonable time frame.

However, finding the optimal number of master points, their locations and the optimal perturbation in those locations is not always easy. Previously, we have shown that SSC-GA can deliver good results using any randomly selected master point locations, as long as they cover the overall reservoir space[11]. This paper extends that work to show that using a particular linked-list implementation, SSC-GA can identify the number of master points necessary to capture the heterogeneity of the reservoir permeability and deliver good models that match closely to the reference field.

These encouraging results suggest that the SSC-GA technique is potentially capable of delivering good results for real world reservoir fields. We continue this work by developing and incorporating cheap surrogates for reservoir simulator to reduce the modeling time, which is crucial to the success of deploying the SSC-GA system.

Acknowledgment

The GA implementation is based on the public domain library GALIB written by Matthew Wall at MIT.

References

[1] K. A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*, Ph.D. Dissertation, Department of Computer and Communication Science, University of Michigan, Ann Arbor, MI, 1975.

[2] K. A. De Jong and W. M. Spears, ”An analysis of the interacting roles of population size and crossover in genetic algorithms.” *Proc. of Parallel Problem Solving from Nature*, 1990, pp. 38-47.

[3] C. V. Deutsch, *Geostatistical Reservoir Modeling*, Oxford University Press, 2002.

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

[5] J. J. Gomez-Hernandez, A. Sahuquillo and J. E. Capilla “Stochastic simulation of transmissivity fi elds conditional to both transmissivity and piezometric data, 1. The theory” *Journal of Hydrology*, 203(1-4) 1997, p.162-174.

[6] C. He and B.X. Hu ”Using sequential self-calibration and genetic algorithm methods to optimally design tracer test for estimation of conductivity distribution,” *Proc. of the 15th International Conference on Computational Methods in Water Resources (CMWR XV)*, June 13-17, 2004 Chapel Hill, NC, USA

[7] J. Knowles, ”ParGO: a hybrid algorithm with on-line landscape approximation of expensive multiobjective optimization problems,” *IEEE Transactions On Evolutionary Computation*, 10(1):50-66, 2006.

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

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

[10] X-H. Wen, C. V. Deutsch and A. S. Cullick, ”Inversion of dynamic production data for permeability: fast streamline-based computation of sensitivity coeffi cients of fractional flow rate,” *J. of Hydrology,* vol. 281, 2003, pp. 296-312.

[11] X.-H. Wen, T. Yu and S. Lee “Coupling sequential self-calibration and genetic algorithms to integrate dynamic production data in geostatistical modeling,” *Proc. of the Seventh International Geostatistics Congress, 2004*, Banff, Alberta, Canada Sept. 2004, pp. 691–702.

[12] X.-H. Wen, S. Lee and T. Yu “Simultaneous integration of pressure, water cut, and 4-D seismic data in geostatistical reservoir modeling,” *Mathematical Geology*, Volume 28, Number 3, April 2006.

[13] W. W.-G. Yeh ”Review of parameter identifi cation procedure in groundwater hydrology:The inverse problem,” *Water Resource Research*, Volume 22, Number 2, 1986, pp. 85–92.

[14] Z. Z. Zhou, Y. S. Ong, P. B. Nair, A. J. Keane and K. Y. Lum, Combining global and local surrogate models to accelerate evolutionary optimization, *IEEE Transactions On Systems, Man and Cybernetics - Part C*, in press