Scatter Search Metaheuristic Applied to the History-Matching Problem
Sergio H. G. Sousa, SPE, Célio Maschio, SPE, and Denis J. Schiozer, SPE, UNICAMP

Abstract
Reservoir simulation plays an important role in managerial decisions during the life of a reservoir, so it is vital that they present an accurate prospect of real reservoir performance. The main goal of the History Matching (HM) process is to improve the quality of numerical models by constraining simulated to observed data. A typical HM process evaluates a chosen objective function (OF) comprised by linear functions on the uncertain reservoir properties. The OF can be modeled in such a way that the HM process can be solved as an optimization problem. This paper discusses how the Scatter Search (SS) technique can solve the HM problem. The main feature of SS is that it works on a set of solutions called the reference set (RefSet). The idea is to improve the overall quality of the RefSet. New solutions are generated by a non-convex combination of explored solutions. The goal of this paper is solve the HM with SS and to evaluate its performance.

The proposed methodology was tested with two base synthetic reservoir models. The first is a homogenous reservoir with 8 different horizontal permeability regions, while the second is a highly heterogeneous reservoir model where low quality background sand is crossed by high permeability canals. The results show that SS was quite efficient, considering the quality of the generated solutions and the number of required numerical simulations.

Most of the current HM methodologies do not perform well when the solution space is large and complex. The application of the SS methodology to the HM problem is a novel approach. Unlike most metaheuristics, SS can be effective even when the simulation time of each tentative solution to the problem is long.

Optimization Background
This section introduces a few terms and concepts common in the optimization field which are important to better understand this work.
Combinatorial Optimization.
Several optimization problems consist of finding the best configuration of a set of variables in order to achieve a predetermined goal. A common type of optimization problem arises when solutions are encoded with discrete values. In this case, every tentative solution is formed by a combination of discrete values, one for each variable that is part of the solution. This is why such problems are called Combinatorial Optimization problems [2]. An important quality of CO problems is that the solution space, the set of all possible combinations of variables, is finite, although its size is combinatorially large. Practical solution space sizes make it impossible to enumerate all solutions, requiring some kind of search method to find good solutions while evaluating only a small fraction of all possible solutions.

Metaheuristics.
Since the early 80s, CO problems have been increasingly solved by a type of algorithm that is currently known as metaheuristics. Metaheuristics are formed combining a low level heuristic with a higher level framework that guides the search for improved solutions [2]. Some examples of metaheuristic methods are Genetic Algorithms (GA), Simulated Annealing, Tabu Search (TS), and Scatter Search. The latter was used in this work to solve the HM problem.

Scatter Search (SS).
The SS methodology is a hybrid metaheuristic method that combines some elements of Evolutionary Algorithms (EA), e.g. GA, and other elements from the TS methodology.

From EAs, SS inherits the concept of population, the Reference Set (RefSet) in SS terms, and the reproduction scheme, where two or more solutions from the RefSet are combined to generate new trial solutions. The main differences are that the RefSet contains much less solutions than EA populations, and the solution combination does not randomly mix elements from two solutions; it prefers to generate new trial solutions via a non convex linear combination of reference solutions (solutions present in the RefSet).

From TS, the SS methodology adopts the tabu list concept, which is a list of solutions that should not be used to generate new trial solutions for the sake of diversity. A solution remains in the tabu list for a fixed number of iterations known as the tabu duration.

At each iteration, two or more reference solutions are combined to generate a set of trial solutions. After these are evaluated according to a specified Objective Function (OF), some (or all) solutions are optimized via a local heuristic. The optimized solutions are then inserted in the RefSet. A detailed description of the SS methodology is available in a work by Fred Glover [3].

Simulation Optimization.
Simulation and optimization technologies have experimented rapid growth in recent years. The combination of both methods has developed a field of its own called Simulation Optimization (SO). In this arena, “the simulation model is a function (whose explicit form is unknown) that evaluates the merit of a set of specifications, typically represented as a set of values” [4].

This template was adapted to solve the HM problem by using the reservoir simulator as a black box where reservoir parameter values are the input and the resulting simulated reservoir performance is the output. A solver module was placed between the metaheuristic optimizer and the reservoir simulator (Figure 1) in order to relieve the optimizer from the tasks of generating trial reservoir models and processing the simulated data to compute the OF value in order to rank visited solutions.

Direct Search.
The term direct search was first used by Hooke and Jeeves in 1961 and is explained by the authors as: “a sequential examination of trial solutions involving comparison of each trial solution with the ‘best’ obtained up to that time together with a strategy for determining (as a function of earlier results) what the next trial solution will be” [5].

This work uses a direct search algorithm based on the ideas laid out by Hooke and Jeeves as a local heuristic used to optimize solutions generated by the SS methodology.

Methodology
The main goal of the research was to solve a HM problem with the SS metaheuristic. An important step in that direction was the modeling of the HM problem as a CO one. In order to do so, the reservoir model parameter ranges were discretized so that the total number of combinations of the parameter values becomes finite. It was also necessary to define an OF so different combinations of parameter values (trial solutions) could be compared. The chosen OF (Equation 1) is a weighted average of normalized distances of the observed data to the simulated data.

$$OF = \frac{\sum_{j=1}^{n} (w_j \times |\bar{D}_j|)}{\sum_{j=1}^{n} w_j} \times \frac{\sum_{j=1}^{n} (w_j \times |\bar{D}_j|)}{\sum_{j=1}^{n} w_j}$$

(1)

The normalization consisted on dividing all parameter distances, calculated as shown on Equation 2, by the corresponding distance of the base case, which is the source of all trial solutions.

$$D = \frac{\sum_{i=1}^{n} (d_{sim} - d_{obs})}{\sum_{i=1}^{n} (d_{sim} - d_{obs})^2} \times \frac{1}{\sum_{i=1}^{n} (d_{sim} - d_{obs})^2}$$

(2)

The normalization of a single parameter is shown on Equation 3. In a typical HM project, the base case has it’s OF value equal to 1 (one) while improved solutions have fractional OF values.
After the solution space was made combinatorial and the OF was able to compare distinct trial solutions, it was possible to implement and test various SS method variations. The next section describes the best of such variations.

The Scatter Search Implementation.

A general outline of the proposed methodology follows:

**Step 1:** Generation of the initial RefSet. Usually the reference RefSet contains only optimized solutions, but since optimizing solutions is so expensive in the HM domain (each evaluation of a trial solution requires a flow simulation), the initial reference set is generated from an initial (non random) sampling of the solution space. The trial solutions are then ranked according to their OF values and the 5 best become the initial RefSet.

**Step 2:** Trial solution generation. Due to the insertion of non optimized solutions on Step 1, the first iteration is composed solely by the local optimization of the best reference solution. On the remaining iterations, the two best non-tabu solutions are linearly combined in a non convex way to generate 4 new trial solutions. The better of the two solutions is marked as tabu for the next 4 iterations.

**Step 3:** RefSet candidate selection. The 4 new trial solutions are evaluated and ranked. Since the local optimization is expensive, only the best trial solution is optimized via the direct search local optimization algorithm. The best in this case is not necessarily the solution with the lowest OF value, but the solution that is the most diversified on the group, i.e., the trial solution whose parameter values most deviate from the reference solutions.

**Step 4:** RefSet update. The improved trial solution is inserted in the RefSet. The RefSet is allowed to hold up to 20 solutions; the worst solutions are trimmed out by the update algorithm. As a matter of fact, the trimming never occurs in the current implementation because there is an upper bound of 8 to the maximum number of iterations which can generate at most 13 reference solutions.

**Step 5:** If no stop condition has been met, skip back to **Step 2**, or else terminate the program. The current stop conditions are:

1. The number of iterations reaches eight,
2. Three consecutive iterations do not insert new solutions in the RefSet, i.e., the optimized trial solution converges three times to a known reference solution.

### Intensification vs. Diversification

An interesting aspect of the proposed methodology is that there is a mix of intensification and diversification strategies. On step 1, the solutions inserted in the RefSet are the 5 best found in the sampling of the solution space; an example of intensification. On step 2, the reference solutions chosen to generate the next set of trial solutions are the ones carrying the best OF values while the criterion to which trial solution will be optimized (step 3) is its diversification level relative to the existing reference solutions. Finally, the tabu list is another form of diversification. It is used to guarantee that a good lower bound solution does not monopolize the trial solution generation.

### SS Test Cases.

Two synthetic reservoir models were developed to test the proposed methodology. The first model, reservoir model A, is a 46×23×5 simulation grid where all parameters are constant throughout the reservoir except for the horizontal and vertical permeabilities. Figure 2 shows how the horizontal permeability varies on each delimited region (marked as Kx through Kx8). There are two five-spot layouts on the reservoir model: (1) Kx2+Kx3+Kx4+Kx6 and (2) Kx3+Kx4+Kx7+Kx8. All historical data was generated with the horizontal permeabilities present in Table 1. In order to simulate data acquisition errors, a background random noise was inserted in all historical data.

### Table 1 – Model A: target horizontal permeability values

<table>
<thead>
<tr>
<th>Region</th>
<th>Permeability (mD)</th>
<th>Parameter Range Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kx1</td>
<td>6257</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx2</td>
<td>6463</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx3</td>
<td>2256</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx4</td>
<td>1008</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx5</td>
<td>4803</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx6</td>
<td>4119</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx7</td>
<td>3277</td>
<td>1000 – 7500</td>
</tr>
<tr>
<td>Kx8</td>
<td>6824</td>
<td>1000 – 7500</td>
</tr>
</tbody>
</table>

The second test case, reservoir model B, has a much more complex geological model. It was generated with geostatistical techniques. It is a synthetic case based on outcrop data, from Brazil and abroad. These data were collected, compiled and treated, both qualitatively and quantitatively, generating a reliable data base of geometrical parameters of depositional elements [6]. The reference model, used to generate the historical data, was composed by 6 layers with different geological properties crossed by high permeability canals over a background region characterized by lower quality sand. The fine simulation grid had 217×275×6 blocks, 12 vertical wells, 7 producers and 5 injectors. This fine model was used to generate the historical data. Subsequently, the historical data was subjected to random noise in order to simulate measurement errors. Finally, the coarse model, with 43×55×6 grid blocks, was obtained from synthetic seismic data, while the premises and geological parameters were different from the fine model. This entire process was devised to add great uncertainty on the canals spatial distribution and properties.

### Results

Reservoir models A and B where subjected to a HM process. Since the complete parameterization for model A was known, this HM was completely automatic. On the other hand, model B did not have a known parameterization so the HM process consisted in an assisted HM process composed of three phases of local optimizations. At the end of each phase, the quality of the match was manually evaluated and a new parameterization was subjected to the SS methodology. The results follow.
Model A.

On the base case, the reservoir model has a constant horizontal permeability of 4500 mD. The chosen parameterization specified fixed ranges (from 1000 mD to 7500 mD) for all permeability regions defined on Figure 2. Each range was then discretized in 27 equally spaced permeability values. With this setup, the solution space is formed by a total of 27^6 combinations, i.e., over 282 billion possible simulation runs. Clearly, a complete enumeration of the solution space is not possible.

The OF was based on the distances of the observed water production rates of all 6 producer wells. Every term was given an equal weight on the global OF expression (Equation 1).

The SS method ran until the maximum number of iterations (eight) was reached. The first six iterations yielded different local optima while the seventh and eighth iteration repeated previous solutions (Figure 3a). As a matter of fact, iterations seven and eight did not generate any additional simulation runs (Figure 3b), what might indicate that some extra diversification is necessary to guarantee the generation of new solutions on these steps. Probably an increase in the tabu duration should do the trick.

In average, each SS iteration consumed about 187 simulation runs. That includes the simulation of the trial solutions (step 2 of the algorithm), and the simulations generated by the local optimization method.

Figures 4a through 4f show how close the found solution vectors mirror the historical data. The departing point of the method was the base simulation run (the blue solid line). The red stars represent the observed water rate on each producer well.

It can be observed that, except for the well PROD5, any of the 6 found solution vectors could be an adequate solution to the HM problem. If well PROD5 is included in the analysis, the 6 found solution vectors could be an adequate solution to the HM problem.

Table 2 shows what horizontal permeabilities were encoded in each solution vector, in decreasing match order. The first line shows the permeabilities used to generate the history file.

<table>
<thead>
<tr>
<th>Kx₁</th>
<th>Kx₂</th>
<th>Kx₃</th>
<th>Kx₄</th>
<th>Kx₅</th>
<th>Kx₆</th>
<th>Kx₇</th>
<th>Kx₈</th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>6257</td>
<td>4663</td>
<td>2256</td>
<td>1008</td>
<td>4803</td>
<td>4119</td>
<td>3277</td>
</tr>
<tr>
<td>1st</td>
<td>6250</td>
<td>5250</td>
<td>3500</td>
<td>1250</td>
<td>4750</td>
<td>3750</td>
<td>4000</td>
</tr>
<tr>
<td>2nd</td>
<td>5500</td>
<td>4750</td>
<td>2750</td>
<td>1250</td>
<td>4500</td>
<td>3750</td>
<td>3750</td>
</tr>
<tr>
<td>3rd</td>
<td>5250</td>
<td>4500</td>
<td>2750</td>
<td>1250</td>
<td>4250</td>
<td>3250</td>
<td>3750</td>
</tr>
<tr>
<td>4th</td>
<td>5000</td>
<td>4500</td>
<td>4000</td>
<td>1500</td>
<td>3750</td>
<td>2000</td>
<td>3750</td>
</tr>
<tr>
<td>5th</td>
<td>4750</td>
<td>2500</td>
<td>5250</td>
<td>1250</td>
<td>3750</td>
<td>3000</td>
<td>2750</td>
</tr>
<tr>
<td>6th</td>
<td>4750</td>
<td>2750</td>
<td>5250</td>
<td>1250</td>
<td>3750</td>
<td>3000</td>
<td>2500</td>
</tr>
</tbody>
</table>

Observe also that Table 2 brings additional important information: updated parameter ranges for the permeability regions. If further optimization is necessary, a tighter and more refined solution space can be derived from the previous solutions. In this sense, the SS method can work as a tool to reduce geological uncertainties.

Model B.

Since reservoir model B did not have a fixed parameterization to work with, a couple of parameterization schemes were tried. Considering the original reservoir dynamic was based on permeability canals, all HM approaches dealt with permeability manipulation.

The first attempt was to make a global transformation manipulating the permeability maps as indicated on Equation 4.

\[ K_{x^*} = A + B \times K_{x_{\text{block}}} \]  \tag{4} 

Since the reservoir consists basically of regions of high permeability (the canals) and low permeability (the background sand), the idea was to use the A coefficient as a way to upgrade the permeability of the background sand without significantly altering the permeability of the canals (A varied from 0 mD to 500 mD). The B coefficient was devised to increase the permeability mostly on the canals, its values varied from 0.3 to 2.5. In other words, the A coefficient acts towards homogenizing the reservoir while the B coefficient increases heterogeneity. This approach did not work. The best solution almost did not improve the base case performance.

After this mishap, the chosen approach was to perform a series of local manipulations to the horizontal permeability in order to provide a better match. The result was a 3 phase process that is summarized below. Each phase starts by establishing a set of parameters, their ranges and discretization, then the SS methodology is applied and the intermediate solutions obtained at the end of the iterations are tested against the local goals of the optimization. As soon as good fit was achieved, the SS methodology is interrupted so that the next phase can begin.

**Phase 1.**

In this phase, the goals were to increase the injectivity of wells INJ1, INJ2, INJ7 and INJ9 and anticipate the water break at producer well PROD3. The parameterization consisted of replacing the permeability on 5 regions on the reservoir to create 4 canals and 1 barrier. The canal parameters had 15 discrete values while the barrier had 11. The size of this solution space was \(15^5 \times 11\), i.e., a little over 556 thousand possible simulation runs.

After the first iteration of the SS methodology, which cost 35 simulation runs, most of the goals have been met. Injector wells INJ1 and INJ9 mirrored observed data, while INJ2 and INJ7 showed significant improvement (green curve on Figures 6a through 6e). Furthermore, producer well PROD3 presented a good match (green curve on Figure 5a).

**Phase 2.**

The goals of this phase were to anticipate the water break at producer well PROD12 and readjust the water production of PROD8. This time, 4 parameters were used, each composed of 15 discrete values. A satisfactory solution was found after the second iteration, at the cost of 48 simulation runs.
The match of PROD12 was significantly improved (magenta curve of Figure 5g) and PROD8's water cut improved (magenta curve of Figure 5e). The INJ2 and INJ7 water injections got a little worse (magenta curves on Figures 6a and 6c), but the remaining injector wells (magenta curves on Figures 6b, 6d, and 6e) kept their adjustments.

**Phase 3.**

The main goal of this phase was to improve the performance of producer well PROD5. The historical data suggests that the water arrives to this well as a water front, but the permeability maps of layers 4 through 6 show it under the influence of at least INJ7 and INJ2, and, possibly, INJ1. Also, the best match up to this moment shows that there is not enough water around PROD5 as well. For these reasons, this phase concentrates on adjusting the influences of these three injectors over PROD5.

The parameterization in this case also yielded 4 parameters with 15 discrete levels each. The first iteration took 19 simulation runs to generate a result (one of the sample solutions was a local optimum) and it was sufficient to reach the goals imposed on producer well PROD5. At this point, injector wells INJ2 and INJ7 remain a little problematic but there was not enough time to continue the phase approach. The final match is shown as the black curves on Figures 5 and 6.

**Conclusions**

The proposed SS methodology showed a good scalability for matches involving several parameters. The local direct search optimization algorithm presented a good convergence, but is still simulation intensive. This fact determined some simplifications on the full SS template algorithm, e.g., the relaxation of optimality for the initial reference solutions, and the optimization of only one of the trial solutions. If a good quality proxy model for the simulator could be obtained with a relatively small number of simulations, a full blown SS implementation could be tried.

Steps like sensitivity analysis, range aperture and the choice of the number of intervals of the discretization can greatly influence the solution space complexity, hence, the algorithm performance and the quality of the solutions.

A grand obstacle in the way of automatic HM is the parameterization stage, where the experience and knowledge of the reservoir engineer cannot be replaced. Because the correct parameterization of model A was previously known, the resulting match was very precise. The HM process of model B was more difficult because only local permeability changes were attempted while several geological properties were uncertain, and the coarse model showed a diverse dynamic from the fine one. This HM was still attempted in order to test the algorithm in a real world application, where time and limited budgets further restrict what can be done.

The main contributions of this methodology are the capacity of dealing with complex solution spaces, the return of a set of solutions (the RefSet) rather than a single one, and the potential of yielding diversified solutions due to the internal diversification strategies.

**References**

Figure 3 – SS performance on reservoir model A

(a) Absolute Objective Function Value

SS Iterations

(b) Number of Simulation Runs

SS Iterations

(c) Water Rate SC (m³/day)

Time (day)

(d) Water Rate SC (m³/day)

Time (day)
Figure 4 – HM of reservoir model A: distribution of solution vectors

(a) Figure 4(a)  
(b) Figure 4(b)  
(c) Figure 4(c)  
(d) Figure 4(d)  
(e) Figure 4(e)  
(f) Figure 4(f)
Figure 5 – Reservoir model B history match of producer wells
Figure 6 – Reservoir model B history match of injector wells