Optimization of Oil Reservoir Models Using Tuned Evolutionary Algorithms and Adaptive Differential Evolution

Claus Aranha*, Ryoji Tanabe‡, Romain Chassagne† and Alex Fukunaga‡
*Graduate School of Systems and Information Engineering, University of Tsukuba, Ibaraki, Japan
Email:caranha@cs.tsukuba.ac.jp
†Institute of Petroleum Engineering, Heriot-Watt University, Edinburgh, Scotland
Email:Romain.Chassagne@pet.hw.ac.uk
‡Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
Email:{fukunaga@idea.c.u-tokyo.ac.jp,rt.ryoji.tanabe@gmail.com}

***Abstract*—In the petroleum industry, accurate oil reservoir models are crucial in the decision making process. One critical step in reservoir modeling is History Matching (HM), where the parameters of a reservoir model are adjusted in order to improve its accuracy and enhance future prediction. Recent works applied evolutionary algorithms (EAs) such as GA, DE and PSO for the HM problem, but they have been limited to classical versions of these algorithms. A significant obstacle to applying EAs to HM is that each call to the fitness function requires an expensive simulation, making it difficult to tune the control parameters for EAs in order to obtain the best performance. We apply and evaluate state-of-the-art, adaptive differential algorithms (SHADE and jDE), as well as non-adaptive evolutionary algorithms (standard DE, PSO) that have been tuned using standard black-box benchmark functions as training instances. Both of these approaches result in significant improvements compared to standard methods in the HM literature. We also apply fitness distance correlation analysis to the search space explored by our algorithms in order to better understand the landscape of the HM problem.**

I. Introduction

The History Matching (HM) problem is of great importance in the Oil Industry. Given a model for an oil reservoir, history matching of this model is the tuning of the parameters of the model such that the output values for the model (for oil, gas, water, etc) match those that have been measured in the past history of the real reservoir. A good History Matching result means that the model for the reservoir will be accurate, and thus useful for informing and guiding future decision making, e.g., when/how to develop an existing oil field. History Matching is essentially a parameter optimization problem. The goal is to find a set of input parameters which results in an accurate forecast model. However, because it is very difficult to predict the influence of any one parameter in the final output, a black box optimization approach is used, where each parameter set is evaluated in a model simulation. Depending on the complexity of the model, this simulation can take hours or even days for complex cases.

For these reasons, there is recent interest in the use of meta heuristic methods for History Matching, such as Differential Evolution (DE) [1], Particle Swarm Optimization (PSO) [2], Ant Colony Optimization (ACO) [3], etc (see section II). However, most of these approaches have two limitations: (1) They use simple variants of the "classical" evolutionary algorithms, and (2) they rarely report enough information to allow for the reproduction and improvement on their results.

To address these issues, in this paper we evaluate and compare DE and PSO, two methods that have shown good results in History Matching in the Oil Engineering literature in recent years, as well as jDE [4] and SHADE [5], two adaptive differential evolution methods. In particular, SHADE is a state-of-the-art algorithm which has been shown to perform very well in the recent CEC2013 black-box optimization competition [6]. We evaluate these algorithms on three field models: two from real world oil fields, and one from a synthetic field.

It is well known that the performance of Evolutionary Algorithms (EAs) is greatly influenced by control parameter settings, and obtaining good performance on a particular problem can require a significant amount of control parameter tuning. The standard approach to control parameter tuning is to run the algorithm on the problem many times using different sets of control parameters, and picking the best-performing set. While this can be automated and made relatively efficient using tools such as SMAC (Surrogate Model based Configurator) [7], the tuning process fundamentally remains the same, except that auto-tuners seek to intelligently sample promising combinations of control parameters.

However, control parameter tuning of EAs for the History Matching problem is difficult because each call to the fitness function (i.e., evaluating a parametrized version of the reservoir simulator) requires an expensive simulation, where each simulation can take hours or even days. To overcome this problem, we tune each candidate EA using the automated tuning tool SMAC [7], where instead of using the actual History Matching problem as the training instance, we use the set of standard CEC2014 black-box benchmark problem instances as the training set and tune the algorithms given the same number of fitness function evaluations that will be used for the History Matching problem. Thus, we can quickly generate a set of tuned control parameter heuristics that can be applied to the History Matching problem. Our results (Section IV) show that this tuning process obtains control parameter sets that significantly outperform the standard control parameters for History Matching suggested in the petroleum engineering literature.

Our comparative evaluation of both the tuned and untuned versions of all algorithms shows that adaptive differential evolution, particularly SHADE, performs well on the History Matching problem. We also use the individuals generated in the comparison experiment to perform an analysis of the search space structure of the HM problem using Fitness Distance Correlation (FDC) [8], [9]. This allows us to better understand the characteristics of HM as a parameter optimization problem, comparing it to optimization benchmarks, and guiding the search for more efficient methods and strategies.

The rest of the paper is structured as follows. We first describe the History Matching problem (Section II), including a review of previous work and a description of our benchmark oil field data. Section III briefly describes the adaptive DE algorithms that are evaluated in this paper. Section IV presents the experimental comparison of tuned and untuned variants of standard algorithms (DE, PSO) and adaptive DE (jDE, SHADE). Section V analyzes the search space of the History Matching problem using fitness distance correlation analysis. Finally, Section VI discusses our results and concludes with directions for future work.

II. The History Matching Problem

The fate of huge investments in the oil and gas industry rests on decision making. Reservoir decision making, such as field development planning, can be improved by having a more accurate forecast of the future behavior of the reservoir. However, because of the difficulty of directly observing a large part of a field's quantitative data, it is hard to calibrate the field model with realistic values regarding the spatial distributions of its properties. Some of these parameters may be "known" from sparse and noisy data obtained from the reservoir, but it is extremely difficult to measure this data in a precise way for every part of the field.

The usual approach is to make a preliminary estimation of field parameters, such as porosity, permeability and so on, and then run the model in a reservoir simulator. The results of this preliminary run seldom match with the data observed from the field, because of the uncertainty in the initial estimates. Consequently, the parameters are iteratively adjusted until the updated model matches the measured data to a sufficient degree of precision. This iterative process is known as "History Matching" (HM), and used to adjust reservoir parameters. The value for these parameters is adjusted based on the output of the simulator. This process of adjusting the model's parameter based on the dynamic observation of change in the output value is an inverse problem, in comparison to the reservoir model simulator, which solves the forward simulation problem.

A recent review on the History Matching has been made by Oliver and Chen [10]. Usually the problem is approached by some sort of optimization method, be it a gradient based method [11], or a non-gradient, such as simulated annealing [12] or evolutionary algorithms [13]. The main challenges of History Matching are its non-linearity, the potentially very large number of variables, and the excessive computation time (which could take days in the largest cases).

A. Problem Description

An instance of the History Matching problem consists of: a model for the studied field, a set of input parameters, and a set of output observations specifying the behavior of the reservoir in time. For each reservoir model, the exact set of output observations may vary, but they generally include oil, gas and water production for each well in the reservoir.

The goal of History Matching is to find an input parameter set that produces outputs from the simulator that agree with the measured data for that particular reservoir. The relevance of an input parameter set is given by the *misfit value*. The misfit is defined as the difference between the measured data for the field, and the observed data from the simulator. For a given parameter set P, a set of outputs O from the simulator over time, and a set of measurements M, the misfit is calculated as

$$\text{Misfit}(P) = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{|\boldsymbol{O}|}\sqrt{(O_j^i - M_j^i)^2} \tag{1}$$

where T is the number of time steps in the simulation, and $|\boldsymbol{O}|$ is the number of outputs (oil, gas, water) for each time step.

The problem includes two kinds of constraints: box constraints and physical constraints. In this work, we deal with the box constraints by limiting all candidate solutions to a (normalized) feasible region [0,1], where 0 is the minimal accepted value for a parameter, and a value of 1 is the maximum accepted value for that parameter. Physical constraints refer to occasions when the simulation does not converge because a particular combination of parameter values is physically infeasible. To deal with these constraints, we abort a simulation run that takes more than a maximum expected amount of time to complete, assuming that the simulation will never converge in this case. The parameter set associated with this run will be given the maximum misfit value (worst utility). Note that the parameters and box constraints used for optimization are a subset of complete set of model parameters. This subset selection is currently made by a petroleum engineer using domain knowledge (increasing the number of parameters being optimized is a direction for future work).

B. Metaheuristics for the History Matching Problem

Metaheuristics (including Evolutionary Algorithms) have been a popular approach for the History Matching problem for the last decade (see [14] for a recent survey). Among recent approaches, Monfared et al. applied a simple GA to a History Matching problem [15]. Xavier et al. [16] also used GA, using real-valued crossover operators. Hajizadeh tested a variety of metaheuristic algorithms in a series of papers, including ACO [17] and DE [18], on the TS2N oil field. He used standard versions of these algorithms, and found that both performed well. Arash [19] also used both standard DE and PSO in a different History Matching problem, and found that DE outperformed PSO. Reynolds [20] studied a BOA-PSO hybrid for the history matching problem. His result showed that PSO by itself outperformed BOA, and slightly underperformed the hybrid on a small synthetic case.

Based on these three previous lines of work, we use PSO and standard DE as baselines for comparison vs. adaptive

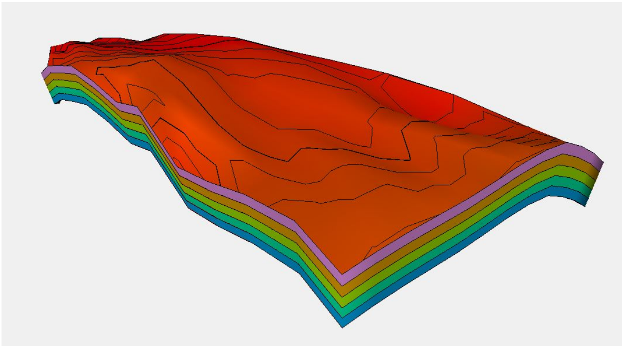


Fig. 1. TS2N field, the colors define its 5 layers

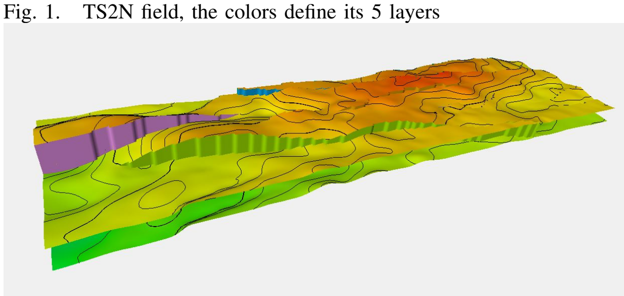


Fig. 2. WATT Synthetic Field

DE variants (jDE and SHADE) in our experiments below. A different, multi-objective approach has been studied by Verga [21] using SPEA2. Multi-objective approaches are of interest for History Matching because for certain decision problems a lower misfit on one set of outputs rather than another may be favored, so a set of diverse solutions may be desirable.

C. Oil Reservoir Models

In this paper we study 3 different reservoir models: the Teal South Reservoir (TS2N), the Northern Sea Field (NSF) and the Watt Field (WATT). The first two models refer to real world fields, while the third one is a synthetic case.

Teal South Reservoir (TS2N): The Teal South reservoir is a real reservoir located in the Gulf of Mexico. Teal South has a single production well that penetrates 4500 ft into the sand. It was initially over-pressured at 3096 psi. Monthly production data of Oil, Gas and Water is listed from 1996 to 1999. There are five geological layers in the model with uniform properties. The TS2N is a real life example, albeit a simple one. A visualization of the TS2N model is shown in figure 1.

WATT Reservoir (WATT):The Watt Field is a semi-synthetic study, which means based in part on real field data and in part on synthetic in order to describe a realistic field example. It has been developed in Arnold [22]. Top structure and wireline data is based on real field data however the fluid properties, relative permeability, and capillary data are synthetic. The field development plan is also synthetic resulting in an artificial production response. The field has been developed through a set of 16 horizontal production wells located across the central part of the reservoir and 5 horizontal and 2 vertical injection wells around the edges. As compared to Teal South this case is much more complex in terms of physics involved. A visualization of the WATT model is shown in figure 2.

TABLE I. DIFFERENCES IN EXPERIMENTAL PARAMETERS FOR EACH MODEL

Parameter	TS2N	WATT	NSF
Input Parameters	13	12	8
Output Parameters	3	8	3
Time per Evaluation	2-10s	1-4h	10-50min

TABLE II. PARAMETERS FOR THE TS2N MODEL

Parameter	Minimum	Maximum
Porosity	0.18	0.377
Water Oil Contract	4880	4935
Rock Compressibility	7.7e-5	9.7e-5
Permeability X (Layer 1)	60	460
Permeability X (Layer 2)	60	460
Permeability X (Layer 3)	60	460
Permeability X (Layer 4)	60	460
Permeability X (Layer 5)	60	460
Permeability Z (Layer 1)	60	460
Permeability Z (Layer 2)	60	460
Permeability Z (Layer 3)	60	460
Permeability Z (Layer 4)	60	460
Permeability Z (Layer 5)	60	460

Northern Sea Field (NSF): The third field is a real reservoir in the North Sea, and is the more complicated among these three cases as it is a complex, real field. Each full field evaluation of the NSF takes 8-10 hours of computing time, so we use an up-scaled (lower resolution) version of the field, following the method described by Christie [23]. Further details about this field are currently confidential.

Table I summarizes the differences between each model, from a simulation point of view: In this table, *Input parameters* is the dimensionality of the optimization problem that we are using in this experiment. The input parameters used for the TS2N and WATT field are described in tables II and III. *Output parameters* is the number of observations that the model outputs at each time step of the simulation, and is used as $|\boldsymbol{O}|$ in equation 1.

Time per evaluation is how long one simulation (one evaluation) of the model would take. This running time is measured on a single core of a computer with an AMD Opteron 2384 CPU and 32GB of RAM.

III. ADAPTIVE EVOLUTIONARY ALGORITHMS

In this paper, we evaluate jDE [4] and SHADE [5]. two state-of-art variants of Differential Evolution (DE) [1]. First, let us briefly review standard DE. A DE population is represented as a set of real parameter vectors $\boldsymbol{x}_i = (x_1, ..., x_D)$, $i = 1, ..., N$, where D is the dimensionality of the target problem, and N is the population size. In each generation t, a mutant vector $\boldsymbol{v}_{i,t}$ is generated from an existing population

TABLE III. PARAMETERS FOR THE WATT MODEL

Parameter	Minimum	Maximum
Porosity Layers 1-4	0.05	0.1
Porosity Layers 5-8	0.05	0.1
Porosity Layers 9-12	0.05	0.1
Porosity Layers 13-16	0.05	0.1
Porosity Layers 17-20	0.05	0.1
Porosity Layers 21-24	0.05	0.1
Porosity Layers 25-32	0.05	0.1
Porosity Layers 33-40	0.05	0.1
Permeability Multiplier (4 parameters)	0.01	1

member $\boldsymbol{x}_{i,t}$ by applying some mutation strategy. One popular mutation strategy is rand/1, defined as follow:

$$\boldsymbol{v}_{i,t} = \boldsymbol{x}_{r_1,t} + F \cdot (\boldsymbol{x}_{r_2,t} - \boldsymbol{x}_{r_3,t}) \quad (2)$$

The indices r_1, r_2, r_3 are randomly selected from $[1, N]$ such that they differ from each other as well as i. The parameter $F \in [0, 1]$ controls the magnitude of the differential mutation operator. After generating the mutant vector $\boldsymbol{v}_{i,t}$, it is crossed with the parent $\boldsymbol{x}_{i,t}$ in order to generate trial vector $\boldsymbol{u}_{i,t}$. Binomial Crossover, the most commonly used crossover operator in DE, is implemented as follows:

$$\boldsymbol{u}_{j,i,G} = \begin{cases} \boldsymbol{v}_{j,i,G} & \text{if } \mathrm{rand}[0,1) \leq CR \text{ or } j = j_{rand} \\ \boldsymbol{x}_{j,i,G} & \text{otherwise} \end{cases} \quad (3)$$

Where, $\mathrm{rand}[0, 1)$ denotes a uniformly selected random number and j_{rand} is a decision variable index uniformly selected from $[1, D]$. $CR \in [0, 1]$ is the crossover rate. After all of the trial vectors $\boldsymbol{u}_{i,t}, 0 \leq i \leq N$ have been generated, each individual $\boldsymbol{x}_{i,t}$ is compared with its corresponding trial vector $\boldsymbol{u}_{i,t}$, keeping the better vector in the population. Below, we say that a generation of trial vector is *successful* if this replacement occurs. This process is repeated until some termination criterion is met.

The performance of DE depends significantly on the control parameters such as F and CR (defined above). In recent years, there have been much research on *adaptive* DE variants that automatically adapt the control parameter values as the search progresses. In this paper, we focus on jDE and SHADE because jDE is among the most cited adaptive DE algorithms, and SHADE is the highest ranking DE method in the CEC2013 competition on on Real-Parameter Single Objective Optimization [6].

jDE [4] assigns a different set of parameter values F_i and CR_i to each $\boldsymbol{x}_i$, which is used for generating the trial vectors. Initially, the parameters for all individuals $\boldsymbol{x}_i$ are set to $F_i = 0.5$, $CR_i = 0.9$. The control parameter values of the trial vectors are inherited from their parents. However, each parameter is randomly modified (within a pre-specified range) with some probability, and modified parameters are kept for the next generation only when a trial is successful.

SHADE [5] is an adaptive DE which based on JADE [24] algorithm, but uses a historical memory of successful parameter settings based parameter adaptation scheme. Success-history based adaptation uses a historical memory M_{CR}, M_F which stores a set of CR, F values that have performed well in the past, and generate new CR, F pairs by directly sampling the parameter space close to one of these stored pairs.

IV. Comparison of Adaptive DE with Previous Methods

In this section, we evaluate the performance of adaptive DEs (SHADE and jDE) on three instances of the History Matching problem: TS2N, WATT and NSF, described in Section II. We compare these adaptive DEs with standard DE and PSO algorithms that have been previously been evaluated in the petroleum engineering literature (see Section II-B [1]). Our choice of jDE and SHADE as our adaptive differential evolution algorithms is explained in Section III.

These four algorithms are used to generate input parameter configurations for the three fields. Because of the computational cost of one evaluation (up to 50 minutes in the case of NSF, and up to 5 hours in the case of WATT), the evaluation budget for all algorithm runs is set to 500 evaluations. As this is an unusually low number, we perform parameter tuning in each of the four algorithms to compensate for it. Additionally, we evaluate a uniform random sampling of $1,000$ parameter sets from each of the three fields as a control sample.

For DE, PSO, jDE, and SHADE, we evaluate two configurations:

Standard configuration: A parameter configuration taken from recent literature. In the case of DE and PSO, the implementation follow suggestions for configuring these algorithms from recent work in the History Matching literature [17], [20]. jDE and SHADE follow the implementation described in Brest [4] and Tanabe [5], respectively.

Tuned configuration: A parameter configuration tuned for the smaller budget of 500 evaluations used in this experiment. The tuned parameter values for each algorithm were calculated using the Surrogate Model Configuration method (SMAC) [7].

For the exact parameter values in the standard and tuned configuration, see Table IV. In the tables and figures, the tuned variants of the algorithms are denoted by the "tuned" prefix (tunedDE, tunedPSO, tunedjDE and tunedSHADE).

A. Parameter Tuning

As explained in Section I, the standard approach to parameter tuning by repeatedly running each algorithm with various control parameter settings on the History Matching problem is impractical because each call to the History Matching fitness function invovles a very expensive call to the reservoir simulator. We therefore perform parameter tuning using a standard set of black-box benhmark functions as the tuning/training instances, based on the belief that tuning the algorithms to perform well across a broad set of black-box benchmark problems should result in good performance on the actual problem of interest (the HM problem).

All 4 algorithms were tuned using Surrogate Model based Configuration tool, SMAC [7]. An algorithm configurator takes as input an algorithm executable, a formal description of the parameters for the algorithm, and a set of training problem instances. It searches the space of possible parameter values by repeatedly generating a candidate configuration (e.g., by local search) and evaluating the configuration on the set of the training instances (or some intelligently selected subset of training instances). The configuration with highest expected utility on the training set is returned. SMAC can be used to tune real-valued, integer-valued, categorical, and conditional parameters [7].[2]

The evaluation function used by SMAC to assess the quality of a candidate EA configuration was the mean of the difference between the solution found by the EA configuration

[1]We used the standard gbest-model PSO using inertia weight strategy [2].

[2]We used the most recent version of SMAC downloaded from the authors' website, http://www.cs.ubc.ca/labs/beta/Projects/SMAC/.

TABLE IV. FOR PSO, DE, JDE, SHADE, THE DEFAULT CONTROL PARAMETER VALUES, AS WELL AS THE BEST PARAMETERS FOUND BY TUNING THE ALGORITHM WITH SMAC USING 10-DIMENSIONAL CEC2014 BENCHMARK PROBLEMS $F_1 \sim F_{30}$ AS TRAINING PROBLEMS. N/A MEANS THAT ITS CONTROL PARAMETER IS NOT INCLUDED IN THE TUNED PARAMETER SETTINGS.

	Parameters	Range	Default	Tuned
PSO	N	[2, 50]	25	10
	w	[0, 1]	0.729	0.462
	$c1$	[0, 4]	1.494	0.627
	$c2$	[0, 4]	1.494	2.516
DE	N	[6, 50]	25	6
	F	[0.1, 1]	0.5	0.730
	CR	[0, 1]	0.5	0.401
	strategy	see the text	rand/1	current-to-$p$best/1
	p	[0, 0.2]	0.05	0.175
	archive rate	[0, 2]	1.0	1.738
jDE	N	[6, 50]	25	7
	F	[0.1, 1]	0.5	0.859
	CR	[0, 1]	0.5	0.130
	strategy	see the text	rand/1	best/1
	p	[0, 0.2]	0.05	n/a
	archive rate	[0, 2]	1.0	n/a
	τ_F	[0.05, 0.3]	0.1	0.063
	τ_{CR}	[0.05, 0.3]	0.1	0.098
SHADE	N	[6, 50]	25	8
	F	[0.1, 1]	0.5	0.856
	CR	[0, 1]	0.5	0.224
	strategy	see the text	current-to-$p$best/1	“
	p	[0, 0.2]	0.05	0.017
	archive rate	[0, 2]	1.0	0.996
	memory size	[1, 20]	10	9

and the optimal value for each benchmark function in the training set, consisting of 30 functions $F_1 \sim F_{30}$ in 10 dimensions from the CEC2014 benchmarks [25]. The maximum number of fitness evaluations for EA was set to 500, following the evaluation budget of the HM optimization problem. Each run of SMAC was limited to $3,000$ EA configurations. For each EA, for each training scenario, SMAC was run 5 times, and we selected the best result out of these 5 runs.

For each EA , the default control parameter values (from [2], [24], [5]), the ranges for the parameters, as well as the values found by SMAC, are shown in Table IV. Binomial crossover was used for all DE variants. For standard DE, 7 possible mutation strategies (rand/1, rand/2, best/1, best/2, current-to-best/1, current-to-best/2, and current-to-$p$best/1 with archive) could be selected.[3] The “current-to-$p$best/1 with archive” strategy has control parameter p and archive rate; these are modeled as conditional parameters [7] in SMAC. The archive size $|\boldsymbol{A}|$ is set to $\mathrm{round}(\text{archive rate} \times N)$.

B. Results

The results of executing the four algorithms (and their tuned variants) in the three History Matching fields are summarized in Table V and Figure 3. Table V shows the misfit value obtained by each method for each field. In the case of the TS2N and NSF fields, these values are the mean of 20 and 10 runs, respectively. The standard deviation of this mean is given between parenthesis. The last column lists the best misfit obtained from 1000 parameter sets drawn from a uniform distribution, as a control. Figures 3a and 3b show box plots of the misfit obtained for each repetition in the TS2N and NSF, respectively.

In the TS2N reservoir, the tuned versions of the algorithms easily outperformed the standard versions. TunedSHADE performed best, slightly ahead of tunedDE (Wilcoxon test $p = 0.10$). Note that tunedDE included not only different values for CR and F, but also uses the current-to-$p$best/1 mutation, which uses a “memory” of successful individuals.

For the NSF field, we observe somewhat similar results, with the tuned versions of DE, jDE and SHADE outperforming standard DE and PSO. In this field, the difference between these three tuned algorithms is unclear (Wilcoxon test $p = 0.48$ for the null hypothesis that tunedDE equals tunedSHADE). For this field, tuned PSO exhibits extremely anomalous behavior, which reason we have not been able to identify.

For the WATT field, because of its extremely long evaluation time, we could only perform a single repetition of the optimization. We observe that this result generally agrees to the other two fields: The optimization methods outperform the reference random search, and tunedDE,tunedjDE,SHADE and tunedSHADE present the lowest misfit.

Overall, on these History Matching instances we observe the following:

- All algorithms performed better than the trivial control algorithm (random sampling).
- The tuned versions of the algorithms generally performed better than the untuned versions, indicating that standard benchmark problems are useful as training problems for tuning evolutionary algorithms for the History Matching problem.
- Among untuned algorithms, SHADE, an adaptive DE consistently performs the best.
- All of the tuned algorithms achieve roughly similar average performance on these 3 instances. However, the stability of the tuned algorithms differ significantly. Among tuned algorithms, tuned SHADE appears to have the most stable performance.

V. ANALYSIS OF THE HISTORY MATCHING PROBLEM SEARCH SPACE USING FITNESS DISTANCE CORRELATION

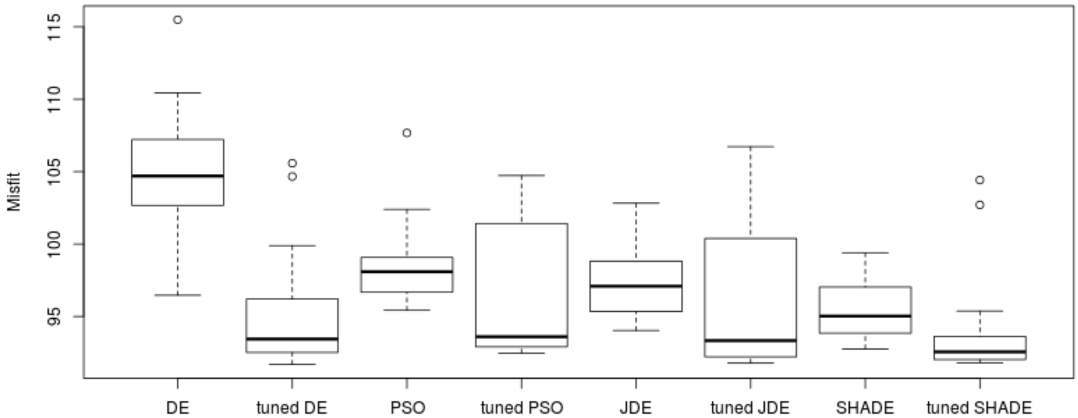
In this section, we experimentally analyze the structure of the search space of the History Matching problem using Fitness Distance Correlation (FDC) [8], [9]. FDC is a landscape analysis method which analyzes the correlation between $f(\boldsymbol{x})$, the fitness of a candidate vector $\boldsymbol{x}$ and $d(\boldsymbol{x}, \boldsymbol{x}^*)$ the distance between $\boldsymbol{x}$ and the optimal solution vector $\boldsymbol{x}^*$. Sampling and plotting $\{f(\boldsymbol{x}), d(\boldsymbol{x}, \boldsymbol{x}^*)\}$ for many points in the search space (e.g., Figure 4) can result in an informative, visualization of the structure of the search space.

EA are often applied to black-box optimization problems such as the History Matching problem where the structure of

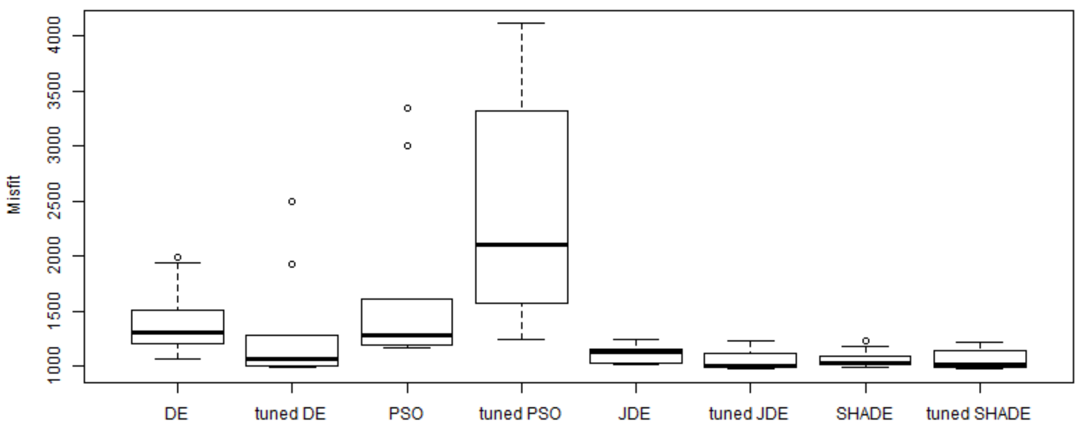
[3] For details on these mutation strategies, see the survey by [26].

TABLE V. MEAN MISFIT VALUES FOR EACH METHOD AND EACH FIELD (STANDARD DEVIATION IN PARENTHESIS)

Field	DE	Tuned DE	PSO	Tuned PSO	jDE	Tuned jDE	Shade	Tuned Shade	Random best
TS2N	104.84(4.66)	95.09 (4.01)	98.49 (2.82)	96.70 (4.69)	97.42 (2.59)	96.17 (5.18)	95.66 (2.2)	93.83 (3.46)	108.62
NSF	1417.84 (318)	1301.33 (509)	1660.24 (812)	2445.19 (1087)	1113.84 (70)	1049.35 (85)	1067.24 (78)	1059.97 (86)	2916.02
WATT	16938.96	16471.63	16510.02	16472.89	16564.55	16469.51	16469.51	16469.06	17490.13



(a) TS2N Model



(b) NSF Model

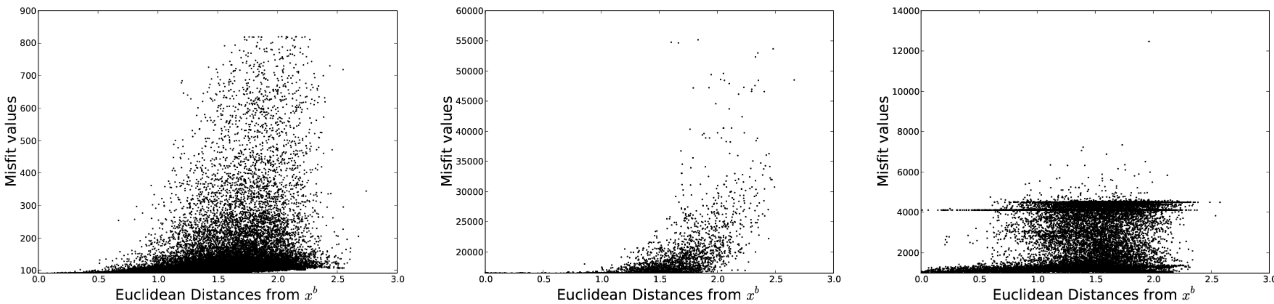
Fig. 3. Box plot of the best misfit values for all repetitions of the TS2N and NSF models. This means 20 repetitions for the TS2N case, and 10 for the NSF case. The vertical axis in each figure is the misfit.

the search space is unknown *a priori*. However, if it is possible to know/predict the structure of the search space a priori, it is possible to adapt the EA to that structure in order to search efficiently. For example, the population could be set to be small or large depending on whether the search space is unimodal or multimodal, respectively. Instances of a class of problems tend to have similar search space structures. Thus, a FDC analysis can guide the development of search algorithms for the History Matching problem. In addition, while there have been a number of landscape analysis on synthetic benchmark functions (c.f. [27]), landscape analysis of real world problems are rare. Thus, we believe that a landscape analysis of the History Matching problem is a valuable contribution.

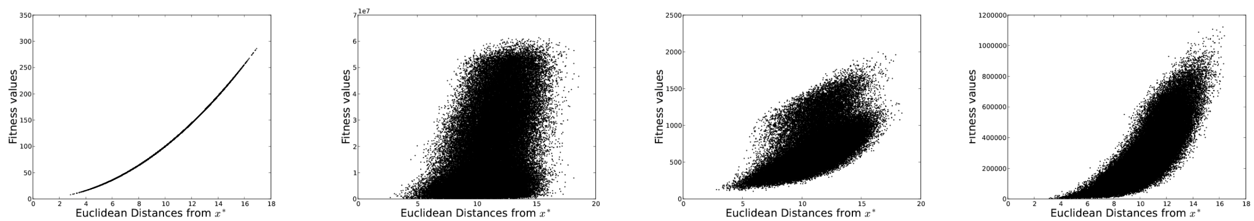
We measured the FDC of the 3 reservoir model instances (TS2N, WATT and NSF) described in Section II. Figures 4(a)–(c) plot $\{f(\boldsymbol{x}), d(\boldsymbol{x}, \boldsymbol{x}^*)\}$ for all candidate vectors $\boldsymbol{X}$ that were evaluated (in all trials, by all algorithms) for each instance in the experiments in Section IV. Since the true optimal solution $\boldsymbol{x}^*$ is unknown, $\boldsymbol{x}^b$, the candidate vector with the lowest (best) fitness value that was found among all of the sampled points, is used as an approximation/proxy for $\boldsymbol{x}^*$, i.e., the fitness distance analysis is performed relative to a point that is a possible, local optimum, instead of the true global optimum.

For comparison and reference, Figures 4(d)–(g) show the FDC for the 10-dimensional Sphere, Ellipsoidal, Rastrigin, and Rosenbrock functions, where the points in the FDC plots were obtained by uniformly sampling the search space 10^5 times. The Sphere and Ellipsoidal functions are unimodal, while the Rastrigin and Rosenbrock functions are multimodal. Also, the Ellipsoidal function is ill-scaled, while the Rosenbrock function is non separable. Detailed definitions and analysis of the properties of the standard functions can be found in [28]. Table VI shows the r_{FDC} metric [9], computed using the following formula:

$$r_{FDC} = \frac{c_{FD}}{s_F s_D} \tag{4}$$



(a) TS2N (b) WATT (c) NSF



(d) Sphere ($D = 10$) (e) Ellipsoidal ($D = 10$) (f) Rastrigin ($D = 10$) (g) Rosenbrock ($D = 10$)

Fig. 4. Figure (a)–(c) plots the Fitness Distance Correlation (FDC) for the three History Matching instances (TS2N, WATT and NSF). All candidate vectors $\boldsymbol{x}$ evaluated by all algorithms, all trials for the experiments in Section IV are represented by a point in the plots. The horizontal axis represents the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{x}^b$, the best vector found for TS2N in Section IV (among all trials of all algorithms), and the vertical axis represents the fitness value of $\boldsymbol{x}$. Figure (d)–(g) show the FDC for the 10-dimensional Sphere, Ellipsoidal, Rastrigin, and Rosenbrock functions, where the points in the FDC plots were obtained by uniformly sampling the search space 10^5 times.

$$c_{FD} = \frac{1}{n}\sum_{i=1}^{n}(f_i - \bar{f})(d_i - \bar{d}) \quad (5)$$

where $n = |\boldsymbol{X}|$, f_i, d_i are the fitness and distance (from $\boldsymbol{x}^b$ or $\boldsymbol{x}^*$) of the i-th vector ($1 \leq i \leq n$), and $\bar{f}, \bar{d}, s_F, s_D$ are the means and standard deviations of fitness values and distances, respectively. As r_{FDC} approaches 1, the search space has a more pronounced big valley structure [8], which is well suited for Evolutionary Algorithms. On the other hand, r_{FDC} close to 0 indicates that the global structure is a weak "needle in a haystack", and a negative value of r_{FDC} indicates a deceptive structure (e.g., multi-funnel landscape).

Figure 4 shows that the FDC plot of the History Matching instances (TS2N, WATT, NSF) resembles that of the standard Ellipsoidal benchmark function. In addition, Table VI shows that for both the History Matching instances and the Ellipsoidal function, r_{FDC} is low. Thus, it can be inferred that TS2N has a search space structure that is somewhat similar to that of the Ellipsoidal function. This result may have a future application: In general, it is desirable for the training used for parameter tuning to be similar to the test problems for which the tuning is ultimately intended. In Section IV-A, our parameter tuning used all 30 instances of the CEC2014 benchmarks [25] as the training set. By using a different set of training instances which focus on problems similar to the Ellipsoid function, we may be able to find control parameter values for EAs that are better than the parameters we found in Section IV-A.

TABLE VI. THE r_{FDC} METRICS [9] FOR EACH FUNCTION.

TS2N	WATT	NSF	Sphere	Ellipsoidal	Rastrigin	Rosenbrock
0.33	0.49	0.22	0.99	0.40	0.61	0.82

VI. DISCUSSION AND CONCLUSION

We applied two adaptive differential evolution algorithms, jDE and SHADE, to three different Oil Reservoir History matching problem instances. We experimentally compared tuned and standard configurations of these adaptive algorithms, as well as non-adaptive standard metaheuristics that have previously used for History Matching (PSO and DE). Our results show that the tuned algorithms significantly outperformed the standard versions from the petroleum engineering literature. Among the untuned algorithms, SHADE consistently performed best. Although all of the tuned algorithms performed comparably, the tuned version of SHADE had the most stable performance.

Standard control parameter tuning of Evolutionary Algorithms is impractical on the History Matching problem due to the extremely expensive fitness function. However, we showed that tuning the EAs using standard black-box benchmark problems as a surrogate (which is orders of magnitude faster than the oil reservoir simulation) is successful. All of the algorithms tuned using this method perform significantly better than their default configurations. In particular, it is worth noting that standard DE and PSO using parameters tuned with this methodology significantly outperformed DE and PSO using control parameter values suggested by previous works on the History Matching problem.

We also analyzed the search space of the History Matching instances using the candidate solution vectors generated during our optimization runs. The TS2N, WATT and NSF models were all shown to have similar search spaces that are visibly similar to the standard Ellipsoidal benchmark function, and they also have r_{FDC} values which are similar to the Ellipsoidal function, suggesting that future work should focus on training optimization algorithms for the HM problem using functions similar to Ellipsoidal function as a surrogate model.

This paper focused on the application of relatively simple, evolutionary algorithms to the HM problem. Application of more complex methods that use surrogate models (e.g., HCMA [29]) is a direction for future work. Another interesting challenge is the addition of seismic constraints to both the model and input parameters. This means using a combination of measured data from the wells (as in this study) with a 3D image from the ground. This might generate an even more precise model, in exchange of a more difficult evaluation function.

Acknowledgment

Thanks to the sponsors of the Edinburgh Time-Lapse Project, Phase V (BG, BP, CGG, Chevron, ConocoPhillips, ENI, ExxonMobil, Hess, Ikon Science, Landmark, Maersk, Nexen, Norsar, Petoro, Petrobras, RSI, Shell, Statoil, Suncor and Total) for supporting Romain Chassagne's work. Thanks to Schlumberger for use of the ECLIPSE™industry standard reservoir simulator software.

Appendix A: Code and Data Availability

All the source code, parameter files and the data for the TS2N and WATT fields, used in this paper, are available at `http://github.com/caranha/HistoryMatchingProject`. The data for the NSF reservoir is not publicly available. To perform the fitness evaluation, a reservoir simulator is required. In the experiments reported in this paper, the ECLIPSE™reservoir simulator was used. The use of other simulation software is possible, but may require some adjustment of the available code.

References

[1] R. Storn and K. Price, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," *J. Global Optimiz.*, vol. 11, no. 4, pp. 341–359, 1997.

[2] Y. Shi and R. E. and, "A modified particle swarm optimizer," in *IEEE CEC*, 1998, pp. 69–73.

[3] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, vol. 26, no. 1, pp. 29–41, 1996.

[4] J. Brest, S. Greiner, B. Bošković, M. Mernik, and V. Žumer, "Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems," *IEEE Tran. Evol. Comput.*, vol. 10, no. 6, pp. 646–657, 2006.

[5] R. Tanabe and A. Fukunaga, "Success-History Based Parameter Adaptation for Differential Evolution," in *IEEE CEC*, 2013, pp. 71–78.

[6] CEC2013, "http://www.ntu.edu.sg/home/EPNSugan/index_files/CEC2-013/CEC2013.htm."

[7] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. of LION-5*, 2011, p. 507523.

[8] K. D. Boese, A. B. Kahng, and S. Muddu, "A new adaptive multi-start technique for combinatorial global optimizations," *Oper. Res. Lett.*, vol. 16, no. 2, pp. 101–113, 1994.

[9] T. Jones and S. Forrest, "Fitness distance correlation as a measure of problem difficulty for genetic algorithms," in *ICGA*, 1995, pp. 184–192.

[10] D. S. Oliver and Y. Chen, "Recent progress on reservoir history matching: A review," *Computational Geosciences*, vol. 15, pp. 185–221, 2010.

[11] F. Zhang and A. Reynolds, "Optimization algorithms for automatic history matching of production data," in *Proceedings of 8th European Conference on the Mathematics of Oil Recovery*, 2002.

[12] M. Sen and P. Stoffa, "Global optimization methods in geophysical inversion," Elsevier Science, 1995.

[13] R. W. Schulze-Riegert, J. K. Axmann, O. Haase, D. T. Rian, and Y. L. You, "Evolutionary algorithms applied to history matching of complex reservoirs," *SPE Reservoir Evaluation and Engineering*, vol. 5, no. 2, pp. 163–172, 2002.

[14] B. Davies, "Limits of automatic history matching," ser. Workshop on conditioning Reservoir Models to Dynamic Data. Barcelona, Spain: EAGE/EUROPEC Meeting, June 2010.

[15] A. D. Monfared and A. Helalizadeh, "Automatic history matching using the integration of response surface modeling with a genetic algorithm," *Petroleum Science and Technology*, vol. 30, no. 4, pp. 360–374, 2012.

[16] C. R. Xavier, E. P. dos Santos, V. da Fonseca Vieira, and R. W. dos Santos, "Genetic algorithm for the history matching problem," *Procedia Computer Science*, vol. 18, pp. 946–955, 2013.

[17] Y. Hajizadeh, M. Christie, and V. Demyanov, "History matching with differential evolution approach; a look at new search strategies," in *SPE Annual Conference and Exhibition*, 2010.

[18] ——, "Ant colony optimization for history matching," in *SPE EUROPEC Annual Conference and Exhibition*, 2009.

[19] A. Mirzabozog, L. Nghiem, Z. Chen, and C. Yang, "Differential evolution for assisted history matching process: Sagd case study," in *SPE Heavy Oil Conference*, 2013.

[20] A. P. Reynolds, A. Abdollahzadeh, D. W. Corne, M. Christie, B. Davies, and G. Willians, "A parallel boa-pso hybrid algorithm for history matching," in *IEEE Congress on Evolutionary Computation*, 2011, pp. 894–901.

[21] F. Verga, M. Cancelliere, and D. Viberti, "Improved application of assisted history matching techniques," *Petroleum Science and Engineering*, vol. 109, pp. 327–347, 2013.

[22] D. Arnold, V. Demyanov, D. Tatum, M. Christie, T. Rojas, S. Geiger, and P. Corbett, "Hierarchical benchmark case study for history matching, uncertainty quantification and reservoir characterisation," *Computers and Geosciences*, vol. 50, pp. 4–15, 2013.

[23] M. A. Christie, "Upscaling for reservoir simulation," *Society of Petroleum Engineers*, 1996.

[24] J. Zhang and A. C. Sanderson, "JADE: Adaptive Differential Evolution With Optional External Archive," *IEEE Tran. Evol. Comput.*, vol. 13, no. 5, pp. 945–958, 2009.

[25] J. J. Liang, B. Y. Qu, and P. N. Suganthan, "Problem Definitions and Evaluation Criteria for the CEC 2014 Special Session and Competition on Single Objective Real-Parameter Numerical Optimization," Zhengzhou Univ. and Nanyang Technological Univ., Tech. Rep., 2013.

[26] S. Das and P. N. Suganthan, "Differential Evolution: A Survey of the State-of-the-Art," *IEEE Tran. Evol. Comput.*, vol. 15, no. 1, pp. 4–31, 2011.

[27] C. L. Müller and I. F. Sbalzarini, "Global characterization of the CEC 2005 fitness landscapes using fitness-distance analysis," in *EvoApplications*, 2011, pp. 294–303.

[28] N. Hansen, S. Finck, R. Ros, and A. Auger, "Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions," INRIA, Tech. Rep. RR-6829, 2009, updated February 2010. [Online]. Available: http://coco.gforge.inria.fr/bbob2012-downloads

[29] I. Loshchilov, M. Schoenauer, and M. Sebag, "Bi-population CMA-ES algorithms with surrogate models and line searches," in *GECCO (Companion)*, 2013, pp. 1177–1184.