Local Optima Networks for Assisted Seismic History Matching Problems

Paul Mitchell, Gabriela Ochoa, Yuri Lavinas, Romain Chassagne

Local Optima Networks for Assisted Seismic History Matching Problems

Paul Mitchell[1,2][0000−0001−6379−0888], Gabriela Ochoa[3][0000−0001−7649−5669], Yuri Lavinas[4][0000−0003−2712−5340], and Romain Chassagne[1,5][0000−0001−5870−6098]

[1] Heriot-Watt University, The Avenue, Edinburgh, EH14 4AS, UK
pcm2@hw.ac.uk
[2] TAQA Bratani Limited, TAQA House, Prime Four Business Park, Kingswells, Aberdeen, AB15 8PU, UK
paul.mitchell@taqaglobal.com
[3] University of Stirling, Stirling, FK9 4LA, UK
gabriela.ochoa@stir.ac.uk
[4] University of Tsukuba, 1-1-1 Tennodai Tsukuba, 305-8577, Japan
yclavinas@gmail.com
[5] Now at Bureau de Recherches Géologiques et Minières (BRGM), 3 avenue Claude-Guillemin, BP 36009 45060, Orléans, Cedex 02, France
r.chassagne@brgm.fr

Abstract. Despite over twenty years of research and application, assisted seismic history matching (ASHM) remains a challenging problem for the energy industry. ASHM attempts to optimise the subsurface reservoir model parameters by matching simulated data to the observed production and time-lapse (4D) seismic data, leading to greater confidence in the assimilated models and their predictions. However, ASHM is a difficult and expensive task that has had mixed results in industry, and a new approach to the problem is required. In this work, we examine ASHM from a different perspective by exploring the topology of the optimisation fitness landscape. Many methods for fitness landscape analysis (FLA) have been developed over the past thirty years, but in this work, we extend the use of local optima networks (LONs) to the real-world and computationally expensive ASHM problem. We found that the LONs were different for objective functions based on both production data and time-lapse reservoir maps, and for each dimensionality. Objective functions based on well pressures and oil saturation maps had the highest success rate in finding the global optimum, but the number of suboptimal funnels increased with dimensionality for all objective functions. In contrast, the success rate and strength of the global optima decreased significantly with increasing dimensionality. Our work goes some way to explaining the mixed results of real ASHM problems in industry, and demonstrates the value of fitness landscape analysis for real-world, computationally expensive problems such as ASHM.

Keywords: Assisted seismic history matching · Local optima networks · Fitness landscape analysis

1 Introduction

In the energy industry, oil and gas production [25], CO_2 storage [11], geothermal energy [8], and more recently Hydrogen storage [26] have been exploited over many decades for numerous purposes. However, to simulate and optimise the production and injection of fluids from the reservoirs, detailed computer models of the subsurface geology, or *reservoir models*, are required [19]. These can be challenging to construct and often contain many errors and uncertainties. The main sources of data that are used to define the models are three-dimensional seismic images of the subsurface geology and the in-situ properties of the rocks measured in well bores. However, well-bore data is only measured at sparse locations in the reservoir and contains measurement errors, and seismic data, despite being recorded over the whole field, has weak signals, is noisy, and contains artefacts. The interpretation and integration of the data to construct reservoir models is a highly skilled task, but the resultant models are typically a simplified representation of the real subsurface geology, and have many possible alternatives [2]. Consequently, their predictions and forecasts often have a wide range of uncertainty.

During production, the reservoir model can be calibrated and constrained using reservoir *history matching* [16]. The model parameters are adjusted to achieve a good match between the predicted production data from the model and the measured, or *observed*, production data from the wells. In addition, time-lapse, or 4D, seismic data are sometimes recorded over the field, which provide three-dimensional images, or snapshots, of fluid movement and pressure changes within the reservoir over time [20]. This can be used in *assisted seismic history matching* (ASHM) [12], where the model parameters are automatically adjusted so that the model accurately predicts both the historical production data from the wells and the time-lapse seismic data [17]. The quality of the history-matched models is usually quantitatively assessed by measuring the misfit, or *fitness*, between the modelled and observed data using a predefined objective function [6] and metric [5]. However, since the data are sparse and contain errors [21], and since the initial reservoir models are so uncertain, there is often low confidence in the resulting assimilated models [17]. After more than twenty years of research and application, ASHM remains a challenging and expensive task with mixed success in industry, and a new approach to the problem is required.

In this work, we take a fresh look at ASHM and examine it from the perspective of the fitness landscape. This is the multidimensional surface that defines the misfit between the model and observed data over the entire parameter search-space [22]. We propose that characterising the topology of the fitness landscape before optimisation will inform the problem setup, guide the choice of optimisation strategy, and lead to better assimilated models. We believe that this new approach will result in a deeper understanding of the ASHM problem and will advance the technology in industry.

However, characterising the topology of fitness landscapes is an extremely challenging task. Fitness landscapes are complex, multidimensional surfaces that are difficult to compute and visualise. Many methods for fitness landscape analy-

sis (FLA) have been developed during the past thirty years or more [9], but they have typically been applied to combinatorial problems or continuous problems with analytical solutions. Exploratory landscape analysis (ELA) is a popular approach to FLA [10], but recently, local optima networks (LONs) [13,14], which depict the global structure of fitness landscapes as graphs, has been extended to continuous optimisation problems [1]. LONs use a basin-hopping approach followed by local minimisation to identify funnels and local or global minima within the fitness landscape structure. They have been applied to benchmark problems and, more recently, to real-world problems with analytical functions [3], but they have not been applied to complex real-world problems such as ASHM.

This paper aims to extend the use of LONs to the computationally expensive real-world problem of ASHM, and to explore the structure of ASHM fitness landscapes for the first time. We have implemented the algorithm to run in parallel on a cluster computer, and calculated the LONs for several objective functions using a realistic reservoir model. These are based on both production data from wells and time-lapse reservoir maps extracted from the simulation model, which represent ideal 4D seismic data. Furthermore, we have explored the impact of the number of parameters by computing LONs in four, seven, and ten dimensions.

The paper is structured as follows. In Section 2.1, we define the local optima network model. Then, in section 2.2, we describe the reservoir model used for the experiments and the method used to sample the fitness landscapes. In section 3, we present the ASHM local optima networks for the reservoir model based on four objective functions and in three different dimensions, as well as their network and performance metrics. We discuss the results and their implications for ASHM in Section 4. Finally, we summarise our observations and conclusions in Section 5, as well as our thoughts on future work.

2 Method

2.1 Local Optima Networks

To analyse and visualise the structure of the studied landscapes, we consider the compressed monotonic local optima network (CMLON) model [15,3]. We formalise the notions of fitness landscapes and local optimum in continuous optimisation before defining the CMLON model. We also define the notions of monotonic sequence and funnel, which are relevant to our analysis. Thereafter, we describe the process of sampling and constructing the network models.

Definitions.

Fitness landscape. Is a triplet $(\mathbf{X}, N, f)$ where $\mathbf{X} \in \mathbb{R}^n$ is the set of all real-valued solutions of n dimensions, *i.e.*, the search space; N is a function that assigns to every solution $\mathbf{x} \in \mathbf{X}$ a set of neighbours $N(\mathbf{x})$; and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the fitness function. A potential solution $\mathbf{x}$ is denoted as vector $\mathbf{x} = (x_1., x_2, \ldots, x_n)$,

and the neighbourhood is based on hypercubes. Formally, the neighbourhood of a candidate solution $\mathbf{x}_k$ is defined as, $\mathbf{x}_j \in N(\mathbf{x}_k) \leftrightarrow |x_{ki} - x_{ji}| < s_i, i = \{1, \ldots, n\}$ where $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ is a vector that represents the size of the neighbourhood in all dimensions.

Local optimum. Is a solution $\mathbf{x}^* \in \mathbf{X}$ such that $\forall \mathbf{x} \in N(\mathbf{x}^*)$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$.

Compressed monotonic LON model. In the standard LON model [13], nodes are local optima and edges represent any possible transition among optima with a given perturbation operator. The CMLON model [15,3] is a coarser model that compresses connected local optima at the same (or very similar) fitness values into single nodes, and has edges representing improving transitions only. The purpose of the CMLON model is to study landscapes with neutrality, that measures the complement of the proportion of compressed nodes to the total number of local optima, providing an indication of plateaus within the landscape, as well as to explore the landscape's funnel structure. To define the graph model, we first define the nodes and edges.

Compressed local optimum. A compressed local optimum is a single node that represents a set of connected nodes in the LON model with the same (or very similar) fitness value.

Monotonic perturbation edges. There is a monotonic perturbation edge from a local optimum $\mathbf{l}_1$ to a local optimum $\mathbf{l}_2$, if $\mathbf{l}_2$ can be obtained from a random perturbation of $\mathbf{l}_1$ followed by a local minimisation process, and $f(\mathbf{l}_2) \leq f(\mathbf{l}_1)$. The edge is called monotonic because the transition between two local optima is non-deteriorating. Edges are weighted with the number of times a transition between two local optima occurred.

Compressed monotonic LON. It is the directed graph $CMLON = (CL, CE)$ where nodes are compressed local optima CL, and edges $CE \subset ME$ are aggregated from the monotonic edge set ME by summing up the edge weights.

Monotonic sequence. A monotonic sequence is a path of connected local optima where their fitness values are always decreasing. Every monotonic sequence has a natural end, which represent a funnel bottom, also called sink in graph theory.

Funnel. We can characterise funnels in the CMLON as all the monotonic sequences ending at the same compressed local optimum (funnel bottom or sink).

Sampling and Constructing the Network Models. Our methodology for sampling and constructing the networks is based on the *basin-hopping* algorithm [23]. Basin-hopping is an iterative algorithm, where each iteration is composed of a random perturbation of a candidate solution, followed by a local minimisation process and an acceptance test. Specifically, the CMLON model construction uses a variant called monotonic basin-hopping (MBH) [7] where the acceptance criterion considers only improving solutions.

This is the first attempt to apply local optima networks to a computationally expensive real-world problem, such as ASHM. In order to construct the models,

several basin-hopping runs are conducted. Each run produces a search trajectory, which is recorded and stored as a set of nodes (local minima) and edges (consecutive transitions). Note that different runs can in principle traverse the same nodes and edges, even if they start from different initialisation points. The models are constructed in a post-processing stage, where the trajectories generated by a fixed number of runs (100 in our implementation) are aggregated to contain only unique nodes and edges.

ASHM problems require simulation and evaluation of the objective function to compute each fitness sample. This is computationally expensive and may take several minutes or even hours to run. To ease this, the original LONs algorithm (https://github.com/gabro8a/LONs-Numerical.git) was re-implemented on a multi-node high-performance cluster computer. Each LON takes approximately four days to compute using fifty CPUs (AMD ™Opteron 6348 processors operating at, 1400MHz) for each objective model. Consequently, some parameters have been estimated, and further evaluation is required. For example, the perturbation strength, which is an important parameter [1], was fixed at 2.5% of the normalised search space distance for each parameter, but this should be determined more rigorously in future work. The computational time also increases with dimensionality. Reservoir simulations were performed using the ™Eclipse 100, version 2018.2, black-oil reservoir simulator.

The CMLON network visualisations were generated using the original LONs post-processing software, which is based on the R implementation of the igraph package (https://igraph.org/r/). The parameters for the CMLON visualisations are listed in Table 1. The *Best* value is the minimum fitness of the objective function for each CMLON. The number of iterations was determined by examining the convergence of each run, and was chosen to be one hundred and fifty for all experiments. However, the increase in value of the *best* fitness with dimensionality suggests that more iterations may be needed to improve convergence in higher dimensions. The position threshold, ϵ, defines where two solutions represent the same local optimum, i.e., the difference between each of their components is less than the threshold value, ϵ.

2.2 Reservoir Model and Fitness Computation

The Brugge model, shown in Figure 1, is a full-field synthetic reservoir model created for the Society of Petroleum Engineers (SPE) Applied Technology Workshop (ATW) on production optimisation in 2008 [18]. Twenty oil production wells and ten water injection wells are used to develop the field; the oil producers are placed near the crest of the structure in the oil zone, and the water injectors are sited down-flank in the water zone. Four geological formations were defined within the reservoir (Schelde, Mass, Waal, and Schie), which are populated with realistic petrophysical values for porosity, net-to-gross, and permeability. The model also contains a geological fault near the crest of the structure.

One of the supplied model realisations (FY-SF-KM-1-1) was selected to represent the true reservoir model, which was then simulated to generate the *observed* dataset used for the experiments. This included oil production rates (WOPR)

Table 1: CMLON parameters for the Brugge model SHM fitness landscapes.

Objective	Dimensions	Best	Step size	Iterations	Runs	ϵ
WOPR	4	1.032	2.5%	150	100	0.1
	7	2.277	2.5%	150	100	0.1
	10	20.325	2.5%	150	100	0.1
WBHP	4	1.597	2.5%	150	100	0.1
	7	1.616	2.5%	150	100	0.1
	10	1.678	2.5%	150	100	0.1
SOIL	4	7.410×10^{-10}	2.5%	150	100	1×10^{-6}
	7	9.030×10^{-7}	2.5%	150	100	1×10^{-6}
	10	8.000×10^{-6}	2.5%	150	100	1×10^{-6}
PRESSURE	4	2.390×10^{-3}	2.5%	150	100	0.1
	7	3.550×10^{-2}	2.5%	150	100	0.1
	10	2.490×10^{-1}	2.5%	150	100	0.1

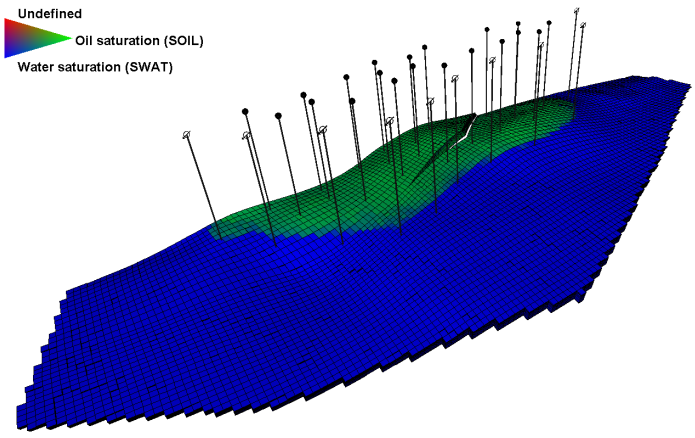


Fig. 1: Three-dimensional view of the Brugge reservoir model. The green zone is the oil bearing reservoir and the blue zone is the water bearing reservoir. The wells are shown by the black lines that penetrate the reservoir from above.

and bottom-hole pressure (WBHP) measurements for each well, and time-lapse reservoir maps (between zero and ten years of production) of reservoir pressure (PRESSURE) and oil saturation (SOIL) extracted from the simulation model. The *observed* production data were used to control subsequent reservoir simulations by total liquid rate, but the proportions of produced oil and water, as well as bottom-hole pressure, were not constrained. These were dependent on the reservoir model and fluid properties.

The reservoir model properties were perturbed to create different realisations of the initial model, and their simulated data were compared with the observed data to calculate the model's fitness. Scalars were applied to the petrophysical property grid values of net-to-gross ratio (NTG) and permeabilities (PERMX and PERMY), measured in millidarcies (mD), within each formation. The parameters and their ranges are listed in Table 2. In the four-dimensional problems, the net-to-gross and permeability values were varied for the Waal formation, as well as the transmissibility of the fault. The same properties were also varied for the Maas formation in the seven-dimensional problems, and, in addition, the same properties were varied for the Schie formation in the ten-dimensional problems.

Table 2: Reservoir model parameters and their geological property ranges. The table shows the mean of the property values for each formation, as well as the minimum and maximum scalar values in parentheses.

Formation	Property	Mean	Minimum	Maximum
Maas	NTG_MAAS	0.88	0.62 (0.7)	0.97 (1.1)
	PERMX_MAAS (mD)	90	45 (0.5)	900 (10.0)
	PERMY_MAAS (mD)	90	45 (0.5)	900 (10.0)
Waal	NTG_WAAL	0.97	0.68 (0.7)	1.0 (1.1)
	PERMX_WAAL (mD)	814	407 (0.5)	1628 (2.0)
	PERMY_WAAL (mD)	814	407 (0.5)	1628 (2.0)
Schie	NTG_SCHIE	0.77	0.39 (0.7)	1.0 (1.3)
	PERMX_SCHIE (mD)	36	18 (0.5)	360 (10.0)
	PERMY_SCHIE (mD)	36	18 (0.5)	360 (10.0)
Fault	FLT_TRANS	1	0.1 (0.1)	2.0 (2.0)

Four objective functions were defined for the experiments, two based on the well production data, and two based on time-lapse reservoir maps extracted from the simulation models. The oil production rates of the wells (WOPR) and bottom-hole pressure measurements (WBHP) were used for the production based objectives, and oil saturation maps (SOIL) and reservoir pressure maps (PRESSURE) were used for the time-lapse reservoir map based objectives. For the well based objectives, fitness values were calculated as the mean square error (MSE) of the model and observed production data for each well and averaged for an overall model fitness. For the map based objectives, the MSE of the modelled and

observed reservoir maps was calculated for each formation, and also averaged for an overall model fitness.

3 Results

The Brugge model CMLONs are shown as two-dimensional networks in Figure 2 and as three-dimensional representations in Figure 3. In the three-dimensional visualisations, fitness is shown on the vertical axis, where lower values are better. In all networks, the pink nodes are global funnels and the blue nodes are suboptimal funnels. The global optima are highlighted by bright red circles, and the suboptimal funnel bottoms are dark blue circles. The size of the nodes is related to the strength of the incoming connections, and the weight of the edges is related to the number of times a transition occurs between two nodes in the sampling process. The CMLONs are shown for four different objective functions; two are based on well production data (WOPR, WBHP), and two are based on time-lapse reservoir maps (SOIL, PRESSURE). The CMLONs dimensionality, n, increases from left to right for each objective function. The network and performance metrics, which are described in Table 3, are shown in Figure 4 and their values are listed in Table 4.

Table 3: Network and performance metric descriptions.

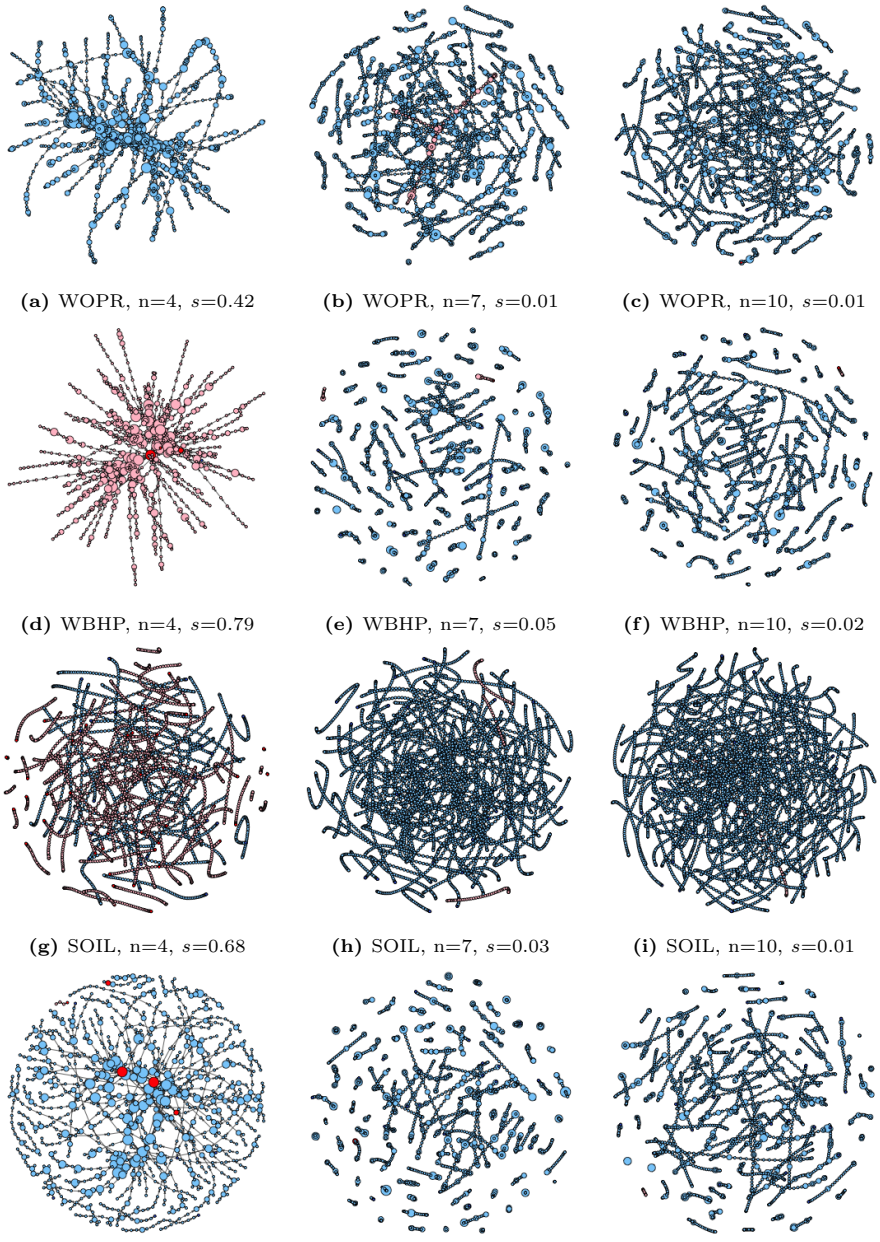
Metric	Description
Success rate	Proportion of basin-hopping runs that reach the global minimum.
Deviation	Mean deviation from the global minimum (unnormalised)
Nodes	Number of nodes in the CMLON (compressed local optima).
Funnels	Number of sinks (CMLON nodes without outgoing edges).
Neutral	Proportion of CMLON nodes to the number of local optima.
Strength	Normalised incoming strength of the globally optimal funnels.

The CMLONs in Figure 2 and their three-dimensional representations in Figure 3 show that the networks have different characteristics for each objective function and for each dimensionality. The four-dimensional CMLON for the WOPR objective has many suboptimal nodes that are clustered toward the centre of the network, and most search trajectories follow a path toward them. In higher dimensions, the funnels are more dispersed and some optimal nodes appear. It has a moderate success rate in four dimensions, which diminishes in higher dimensions. The CMLON for WBHP has high-strength optimal nodes in four-dimensions, which are also clustered toward the centre of the network. The global optimum is reached by many of the search trajectories. It has the highest success rate of all the experiments. In higher dimensions, the search trajectories disperse to form many suboptimal funnels, and the success rate decreases. The

CMLONs for the SOIL objective are denser than the other objectives and have one hundred funnels in all dimensions. In four dimensions, there are many optimal and suboptimal funnels, and the success rate is high; however, the optima are dispersed. In higher dimensions, the trajectories spiral downwards toward suboptimal funnel bottoms, and the success rate diminishes markedly. The four-dimensional CMLON for the PRESSURE objective has a broad distribution of suboptimal nodes and a few global optima. The majority of the search trajectories descend rapidly towards low fitness values and then move towards the centre of the network. In higher dimensions, the suboptimal optimal funnels disperse, and there are only a few global funnels. The success rate is moderate in four dimensions, but also diminishes in higher dimensions.

In four dimensions, there is a distinct difference in character between the well production objectives (WOPR and WBHP) and the time-lapse reservoir map objectives (SOIL and PRESSURE). The optima of the production-based objectives are clustered toward the centre of the CMLON, whereas the optima of the time-lapse reservoir map based objectives are more dispersed. However, the differences are less apparent in higher dimensions. The SOIL objective is distinctive by its high density of funnels in all dimensions, which are dispersed across the network. They also have more evenly distributed fitness, as indicated by Figures 3g to 3i. The WBHP and SOIL objectives have the highest success rates, which suggests they may be preferable for optimisation.

The network and performance metrics are shown as graphs in Figure 4, and are tabulated in Table 4. They show that the total number of optima (optimal and suboptimal) is large for all objective functions, and that they increase with dimensionality for all objectives. There are noticeably more optima for the SOIL objective than the other objectives. The WOPR, SOIL, and PRESSURE objectives have only a few funnels in four dimensions, but increase markedly in higher dimensions. The SOIL objective is the exception and has one hundred funnels for all dimensions. Overall, the success rate is high for the objectives in four dimensions, but decreases considerably in seven and ten dimensions. The WBHP and SOIL objectives have the highest success rate and the highest strength in four dimensions, which is related to their larger proportion of global optima. The strength of the global optima nodes follows a similar trend to the success rate; where the objectives have high strength nodes in four dimensions, but it decreases in higher dimensions. The WOPR objective is an exception because despite its moderately high success rate, it has low strength. This may be explained by its high deviation, indicating the optima are dispersed. The neutral values are quite different for each objective function. WBHP and SOIL have high neutrality in four dimensions, which decreases in higher dimensions for SOIL, but increases slightly for WBHP in seven dimensions before decreasing in ten dimensions. PRESSURE has lower neutrality for all dimensions, which increases slightly in higher dimensions, and WOPR has negligible neutrality for all dimensions. Deviation depends on the measurement unit of the objective function, which varies widely in these experiments. It is lowest in four dimensions



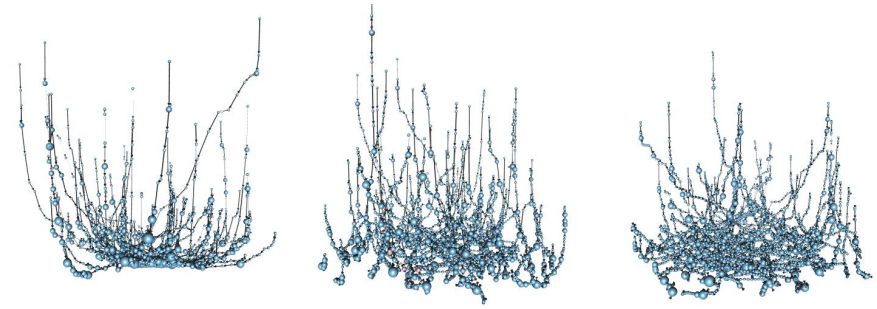
(a) WOPR, n=4, s=0.42 **(b)** WOPR, n=7, s=0.01 **(c)** WOPR, n=10, s=0.01

(d) WBHP, n=4, s=0.79 **(e)** WBHP, n=7, s=0.05 **(f)** WBHP, n=10, s=0.02

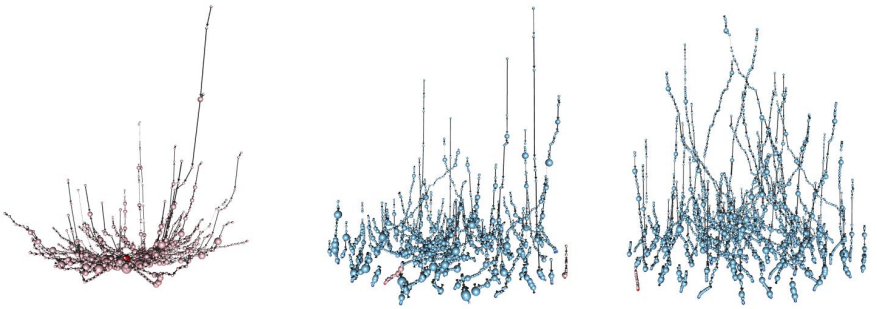
(g) SOIL, n=4, s=0.68 **(h)** SOIL, n=7, s=0.03 **(i)** SOIL, n=10, s=0.01

(j) PRESSURE, n=4, s=0.29 **(k)** PRESSURE, n=7, s=0.01 **(l)** PRESSURE, n=10, s=0.01

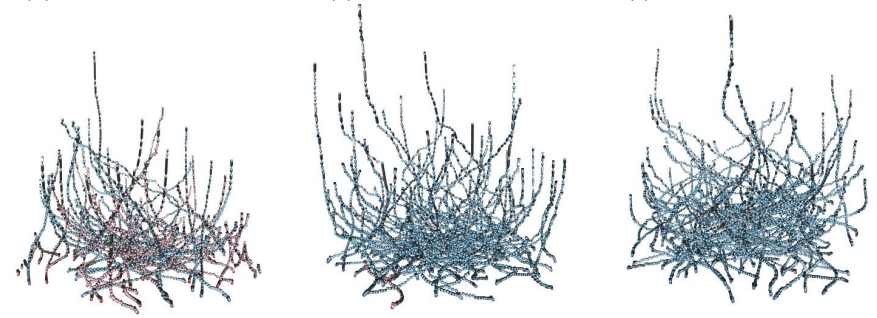
Fig. 2: CMLONs for the Brugge reservoir ASHM fitness landscapes. The CMLONs are shown for four objective functions, and in $n = \{4, 7, 10\}$ dimensions. The success rate, s, is shown for each CMLON.



(a) WOPR, n=4, s=0.42 **(b)** WOPR, n=7, s=0.01 **(c)** WOPR, n=10, s=0.01



(d) WBHP, n=4, s=0.79 **(e)** WBHP, n=7, s=0.05 **(f)** WBHP, n=10, s=0.02



(g) SOIL, n=4, s=0.68 **(h)** SOIL, n=7, s=0.03 **(i)** SOIL, n=10, s=0.01

(j) PRESSURE, n=4, s=0.29 **(k)** PRESSURE, n=7, s=0.01 **(l)** PRESSURE, n=10, s=0.01

Fig. 3: Three-dimensional CMLONs for the Brugge reservoir ASHM fitness landscapes. The vertical axis represents fitness. The CMLONs are shown for four objective functions, and in $n = \{4, 7, 10\}$ dimensions. The success rate, s, is shown for each CMLON.

for all the objective functions, but increases with dimensionality. This follows because the funnels become more dispersed in higher dimensions.

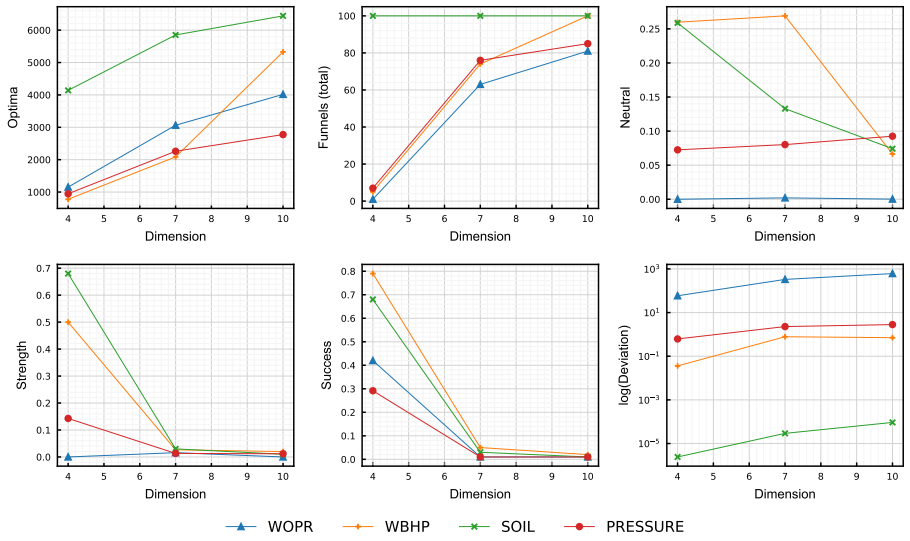


Fig. 4: CMLOM network and performance metrics for the ASHM fitness landscapes.

4 Discussion

The CMLONs presented in Section 3 provide some insight into the structure of the Brugge model ASHM fitness landscapes. The high proportion of suboptimal funnels suggests that the landscapes are complex surfaces with many local minima, even in low-dimensional spaces. This suggests that there may be many non-unique solutions, which is consistent with the results seen in industry [17], and may explain the low-confidence in the models. The number of suboptimal funnels increases in higher dimensions, and they spread out, indicating that there is a reduced likelihood of locating the global optimum. We have explored the problems for four, seven, and ten dimensions; however, ASHM problems for real reservoir models typically have hundreds or thousands of uncertain parameters, or dimensions, which makes them more challenging. The network and performance metrics show that problem difficulty and non-uniqueness increase with dimensionality, by increasing suboptimal funnels and decreasing success rates. It is possible that these trends would extend to higher dimensions, which would exacerbate the problem difficulty, but that requires further investigation.

Four objective functions have been investigated in this work, but which of them should be used and how they should be combined is an open question

Table 4: Network and performance metrics.

Objective	Dimension	Optima	Funnels (opt,subopt)	Neutral	Strength	Success	Deviation
WOPR	4	1153	(0,1)	0.000	0.000	0.420	5.88×10^{1}
	7	3063	(1,62)	0.002	0.016	0.010	3.31×10^{2}
	10	4016	(0,81)	0.000	0.000	0.010	6.12×10^{2}
WBHP	4	778	(2,3)	0.260	0.500	0.790	3.60×10^{-2}
	7	2086	(2,72)	0.269	0.027	0.050	7.81×10^{-1}
	10	5328	(2,98)	0.066	0.020	0.020	7.07×10^{-1}
SOIL	4	4141	(68,32)	0.259	0.680	0.680	2.42×10^{-6}
	7	5851	(3,97)	0.133	0.030	0.030	2.92×10^{-5}
	10	6441	(1,99)	0.074	0.010	0.010	9.26×10^{-5}
PRESSURE	4	947	(1,6)	0.072	0.143	0.292	6.20×10^{-1}
	7	2260	(1,75)	0.080	0.013	0.010	2.28
	10	2775	(1,84)	0.092	0.012	0.010	2.82

[5]. The production-based objectives have slightly higher success rates than the time-lapse reservoir map objectives, although this difference diminishes in higher dimensions. The WBHP and SOIL objectives have the highest success rates, which indicates that both production-based objectives and time-lapse reservoir map objectives are useful in seismic history matching. Combining these may lead to better objective functions with even higher success rates. CMLONs may provide a method for this, potentially leading to improved objective functions with less non-uniqueness and higher success rates.

This work is a first attempt to apply LONs to a computationally expensive real-world problem such as ASHM, and has some limitations. The perturbation step-size was chosen, based on previous experience, to be 2.5% of the normalised ([0,1]) parameter space distance, but it is an important parameter which requires further tuning [1]. The number of iterations was selected by examining the convergence of each run, and balancing it against the computation time. One hundred and fifty iterations was sufficient for the four dimensional problems, but more iterations may be required for higher dimensions. The increased value of the best fitness, the larger proportion of suboptimal funnels, and the reduced success rates seen in these experiments may be related to poorer convergence in higher dimensions; however, significantly increasing the number of iterations could be impractical.

The Brugge model was chosen for these experiments because it is a realistic full-field reservoir model, and its reservoir simulations are very fast. This makes it possible to calculate hundreds of thousands of fitness samples for the LONs. In contrast, reservoir simulations for real producing fields take much longer, and it will be much more challenging to compute LONs in a reasonable time. New methods, such as surrogate or proxy models instead of costly numerical simulations, may be required to significantly speed up simulations and fitness computations

[24]. Time-lapse reservoir maps extracted from the simulation models were used to represent ideal time-lapse seismic data in these experiments. However, real time-lapse seismic data are recorded as acoustic signals, which must be inverted to pressure and saturation maps to compare with reservoir simulation models [4]. Since the signals are weak and the processed data contains noise and other errors, fitness calculations for real data will have some uncertainty. Local optima networks may provide the means to explore the impact of these issues on the fitness landscapes, and their implications for the assimilated models.

The results of this work support our initial hypothesis that characterising the ASHM fitness landscapes will help to understand the problem at a deeper level and guide its definition to achieve more reliable assimilated models. Our work demonstrates the value of analysing fitness landscapes for computationally expensive problems. Local optima networks provide a means to extract key features of the landscape's structure, and network graphs allow them to be visualised intuitively. The results of this work go some way to explaining the mixed results of ASHM seen in industry [17], and offer the potential to help design better objective functions for improved models and more confident predictions.

5 Conclusion

In this work, we have extended the application of local optima networks (LONs) to the computationally expensive, real-world optimisation problem of assisted seismic history matching (ASHM). We have compared the compressed monotonic LONs (CMLONs) of four different objective functions based on both well production data and time-lapse reservoir maps and in three different dimensions.

We found that the CMLONs of the ASHM landscapes have different characteristics for each of the objective functions and dimensionalities. They typically have a few global funnels in four dimensions, but are dominated by a larger number of suboptimal funnels in higher dimensions, where there are few if any global funnels. This implies that there is only a small likelihood of finding the global optimum in higher dimensional problems, which may help to explain the uncertainty in the results of real ASHM problems. Furthermore, it demonstrates the benefit of characterising fitness landscapes before optimisation, even for computationally expensive problems, and may provide a means to design more successful objective functions.

This work represents a first attempt to characterise the fitness landscapes of real-world ASHM problems, but there are many unresolved issues to address. At this stage of research, our goal is to identify methods that can characterise the main features of the fitness landscape and provide a more in-depth understanding of the ASHM optimisation problem. Local optima networks provide a unique view on the global structure of the landscape, but are computationally expensive. This work has investigated relatively low-dimensional problems, but real ASHM problems will have many more parameters.

The Brugge model was selected for these experiments because it is a realistic reservoir model, with a known solution, and fast reservoir simulations. This

allowed for sufficiently well sampled LONs to be computed in a reasonable time. In the future, we intend to further extend the application of LONs to a real producing reservoir with real time-lapse seismic data However, real reservoir models have considerably longer simulations, and it may be more challenging to compute sufficiently sampled LONs. Faster methods for reservoir simulation, such as proxy models, may be required. Furthermore, real time-lapse seismic data contain many sources of data errors, and the impact of these on the fitness landscapes will be considered. Since the true model is unknown for real reservoirs, we will also investigate the impact of model uncertainty on the fitness landscapes and its implications for the assimilated models.

Acknowledgement

We thank the SPECIES society for funding a visiting scholarship for Yuri Lavinas to the University of Stirling, Scotland, UK.

References

1. Adair, J., Ochoa, G., Malan, K.M.: Local optima networks for continuous fitness landscapes. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1407–1414. GECCO '19, Association for Computing Machinery, New York, NY, USA (2019)
2. Arnold, D., Demyanov, V., Tatum, D., Christie, M., Rojas, T., Geiger, S., Corbett, P.: Hierarchical benchmark case study for history matching, uncertainty quantification and reservoir characterisation. Computers & Geosciences **50**, 4–15 (2013)
3. Contreras-Cruz, M.A., Ochoa, G., Ramirez-Paredes, J.P.: Synthetic vs. Real-World Continuous Landscapes: A Local Optima Networks View, pp. 3–16. Springer International Publishing (2020)
4. Corte, G., Dramsch, J., Amini, H., Macbeth, C.: Deep neural network application for 4d seismic inversion to changes in pressure and saturation: Optimising the use of synthetic training datasets. Geophysical Prospecting (2020)
5. Hallam, A., Chassagne, R., Aranha, C., He, Y.: Comparison of map metrics as fitness input for assisted seismic history matching. Journal of Geophysics and Engineering **19**(3), 457–474 (06 2022)
6. He, Y., Aranha, C., Hallam, A., Chassagne, R.: Optimization of subsurface models with multiple criteria using lexicase selection. Operations Research Perspectives **9**, 159–172 (2022)
7. Leary, R.H.: Global optimization on funneling landscapes. Journal of Global Optimization **18**(4), 367–383 (2000)
8. Lund, J.W., Toth, A.N.: Direct utilization of geothermal energy 2020 worldwide review. Geothermics **90**, 101915 (2021)
9. Malan, K.M.: A survey of advances in landscape analysis for optimisation. Algorithms **14**(2), 40 (2021)
10. Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., Rudolph, G.: Exploratory landscape analysis. In: Proceedings of the 13th annual conference on Genetic and evolutionary computation. p. 829–836. Association for Computing Machinery (2011)

11. Michael, K., Golab, A., Shulakova, V., Ennis-King, J., Allinson, G., Sharma, S., Aiken, T.: Geological storage of co2 in saline aquifers-a review of the experience from existing storage operations. International journal of greenhouse gas control **4**(4), 659–667 (2010)
12. Mitchell, P., Chassagne, R.: 4d assisted seismic history matching using a differential evolution algorithm at the harding south field. In: 81st EAGE Conference and Exhibition 2019. vol. 2019, pp. 1–5. European Association of Geoscientists & Engineers (2019)
13. Ochoa, G., Tomassini, M., Vérel, S., Darabos, C.: A study of nk landscapes' basins and local optima networks. In: Proceedings of the 10th annual conference on Genetic and evolutionary computation. pp. 555–562 (2008)
14. Ochoa, G., Veerapen, N.: Mapping the global structure of tsp fitness landscapes. Journal of Heuristics **24**(3), 265–294 (2018)
15. Ochoa, G., Veerapen, N., Daolio, F., Tomassini, M.: Understanding phase transitions with local optima networks: number partitioning as a case study. In: European Conference on Evolutionary Computation in Combinatorial Optimization. pp. 233–248. Springer (2017)
16. Oliver, D.S., Chen, Y.: Recent progress on reservoir history matching: a review. Computational Geosciences **15**(1), 185–221 (2011)
17. Oliver, D.S., Fossum, K., Bhakta, T., Sandø, I., Nævdal, G., Lorentzen, R.J.: 4d seismic history matching. Journal of petroleum science & engineering **207**, 109119 (2021)
18. Peters, L., Arts, R., Brouwer, G., Geel, C., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, N., Vossepoel, F.C., Xu, R., Sarma, P., Alhuthali, A.H.H., Reynolds, A.: Results of the brugge benchmark study for flooding optimization and history matching. SPE Reservoir Evaluation & Engineering **13**(03), 391–405 (2010)
19. Ringrose, P., Bentley, M.: Reservoir Model Design. Springer, 2 edn. (2021)
20. Sambo, C., Iferobia, C.C., Babasafari, A.A., Rezaei, S., Akanni, O.A.: The role of time lapse(4d) seismic technology as reservoir monitoring and surveillance tool: A comprehensive review. Journal of natural gas science and engineering **80**, 103312 (2020)
21. Souza, R., Lumley, D., Shragge, J.: Estimation of reservoir fluid saturation from 4d seismic data: effects of noise on seismic amplitude and impedance attributes. Journal of Geophysics and Engineering **14**(1), 51–68 (2017)
22. Stadler, P.F.: Fitness landscapes. Applied Mathematics and Computation **117**, 187–207 (2002)
23. Wales, D.J., Doye, J.P.: Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. The Journal of Physical Chemistry A **101**(28), 5111–5116 (1997)
24. Werth, B., Pitzer, E., Affenzeller, M.: Surrogate-Assisted Fitness Landscape Analysis for Computationally Expensive Optimization, pp. 247–254. Springer International Publishing (2020)
25. Zhang, G., Qu, H., Chen, G., Zhao, C., Zhang, F., Yang, H., Zhao, Z., Ma, M.: Giant discoveries of oil and gas fields in global deepwaters in the past 40 years and the prospect of exploration. Journal of Natural Gas Geoscience **4**(1), 1–28 (2019)
26. Zivar, D., Kumar, S., Foroozesh, J.: Underground hydrogen storage: A comprehensive review. International journal of hydrogen energy **46**(45), 23436–23462 (2021)