1 **Multi-Solution Well Placement Optimization using Ensemble Learning of Surrogate Models**

2 Mohammad Salehian*, Morteza Haghighat Sefat*, Khafiz Muradov*

3 *Institute of GeoEnergy Engineering, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom*

4   1. Abstract

5   Well location optimization aims to maximize the economic profit of oil and gas field development
6   while respecting various constraints. The limitations of the currently available well placement
7   optimization workflows are their 1) high computational requirements, which makes them
8   inappropriate for full-field applications where a large number of wells have to be optimized using a
9   computationally expensive simulation model; and 2) providing a single optimal solution, whereas on-
10   site operational problems often add unforeseen constraints that result in adjustments to this optimal,
11   inflexible scenario degrading its value.

12   This study presents a multi-solution, surrogate models (SMs)-assisted optimization framework to
13   deliver diverse, close-to-optimum well placement scenarios at a reasonable computational cost.
14   Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm is used as the optimizer while
15   diversity in optimal solutions is achieved by multiple, parallel runs of the optimizer with different
16   starting points. Convolutional Neural Network (CNN) is used as the SM, to partly substitute the
17   computationally expensive reservoir model runs during the optimization process. A new, adjusted
18   Latin Hypercube Sampling (aLHS) procedure is developed to generate initial training datasets with
19   diverse well placement scenarios while respecting reservoir boundaries and well spacing constraints.
20   An ensemble of CNNs is pre-trained using the generated dataset to enhance the robustness of the
21   surrogate modeling as well as to allow estimation of the SM's prediction quality for new data points.
22   The ensemble of CNNs is adaptively updated during the optimization process using selected new data
23   points, to improve the SM's prediction accuracy. To the best of our knowledge, this is the first
24   application of ensemble learning strategy to a well placement optimization problem.

25   The added value of the framework is demonstrated by comparing three optimization approaches on
26   the Brugge and Egg field benchmark case studies. The approaches are 1) 'no SM': using the actual
27   reservoir model only, 2) 'Offline SM': the optimization is performed using SM-only that is pre-trained
28   using initial training datasets generated by the actual reservoir model, and 3) 'Online SM': pre-trained
29   CNNs are adaptively updated during the optimization process using new datasets generated using the
30   actual reservoir model. The surrogate-assisted optimization approach substantially reduced the
31   computation time, while a greater objective value was achieved by employing the adaptive learning
32   strategy due to the enhanced prediction accuracy of the SMs. Multiple diverse solutions were
33   obtained with different well locations but close-to-optimum objective values, which allows a more
34   efficient exploration of the search space at a significantly reduced computational cost. The presented
35   workflow integrates critical challenges that are correlated, yet often addressed independently,
36   providing the much-required operational flexibility and computational efficiency to field operators
37   when selecting from the optimal well placement scenarios.

38   2. Introduction

39   Optimization algorithms are employed to maximize field performance by optimizing one or multiple
40   types of decision variables such as the number, type, or location of new wells (Bangerth et al., 2006;
41   Wang et al., 2012; Al-Ismael et al., 2018; Tavallali et al., 2018; Ding et al., 2019; Jesmani et al., 2020),

42 or the control settings of existing production/injection wells (Haghighat Sefat et al., 2016; Liu and
43 Reynolds, 2016; Wang et al., 2016; Jiang et al., 2019; Arouri and Sayyafzadeh, 2020; Salehian et al.,
44 2020a) while honoring various operational constraints. This, in general, results in a high-dimensional,
45 constrained optimization problem with a computationally expensive objective function based on the
46 simulated reservoir model.

47 Current optimization workflows can be classified into three main groups based on the employed
48 optimization algorithm: (1) adjoint gradient-based algorithms (Van Essen et al., 2011; Kahrobaei et al.,
49 2013; Tavallali et al., 2013; Forouzanfar and Reynolds, 2014; Bukshtynov et al., 2015; Volkov and
50 Bellout, 2018), (2) derivative-free algorithms such as the particle swarm optimization algorithm
51 (Eberhart and Kennedy, 1995; Panahli, 2017; Wang et al., 2019; Ding et al., 2020) or the genetic
52 algorithm (Holland, 1975; Almeida et al., 2010; Lu and Reynolds, 2020; Ma and Leung, 2020), and (3)
53 stochastic approximated gradient-based algorithms such as the Simultaneous Perturbation Stochastic
54 Approximation (SPSA) (Spall, 1992) or the Stochastic Simplex Approximate Gradient (StoSAG) methods
55 (Fonseca et al., 2017; Liu and Reynolds, 2020). The adjoint gradient-based methods are
56 computationally attractive, however, access to the reservoir simulation source code is required to
57 calculate the gradient, which makes them impractical for use with commercial (black box) simulators.
58 The derivative-free algorithms have the advantage of the global search for the optimal solution from
59 all types of decision variables (e.g., categorical, integer, or continuous variables). However, they
60 typically require a large number of function evaluations, and their performance is degraded rapidly
61 with the increasing number of decision variables (Zingg et al., 2008). The approximate gradient-based
62 algorithms overcome the above issues by stochastically estimating the gradient of a black-box
63 objective function using a reasonably sized ensemble (normally containing between 3 to 5 members)
64 of simultaneous perturbations of decision variables. These algorithms have been successfully
65 employed to solve large-scale well placement (e.g. Jesmani et al. (2016) used SPSA), well control (e.g.
66 Haghighat Sefat et al. (2016) used SPSA and Lu et al. (2017) used StoSAG), or multi-level well
67 placement and control problems (e.g. Li et al. (2013) and Salehian et al. (2020b) used SPSA). Salehian
68 et al. (2020a) recently developed a multi-solution optimization framework and showed its value in
69 providing multiple solutions with close-to-optimum objective values. They recommended that the
70 diversity of the provided optimal solutions can be increased by performing multiple parallel
71 optimization runs with different starting points, which however would result in a significantly higher
72 computational cost.

73 Surrogate models (SMs, also known as proxy models) are employed as an approximation method in
74 the optimization process to reduce the cost of objective function evaluations when the underlying full-
75 physics model is expensive to simulate. Three main types of surrogate modeling approaches are
76 commonly employed in the field development and control optimization problems: (1) physics-based
77 approaches such as reduced order modeling (Van Doren et al., 2006; Cardoso and Durlofsky, 2010;
78 Durlofsky, 2010; He and Durlofsky, 2014; Trehan and Durlofsky, 2016) or streamline-based simulation
79 methods (Thiele and Batycky, 2003; Park and Datta-Gupta, 2011; Salehian and Çınar, 2019; Ushmaev
80 et al., 2019), (2) Machine Learning (ML) techniques such as support vector machine (SVM) (Drucker et
81 al., 1997; Guo and Reynolds, 2018; Panja et al., 2018; Zhang et al., 2021), Artificial Neural Network
82 (ANN) (Jain et al., 1996; Güyagüler et al., 2002; Yeten et al., 2003; Golzari et al., 2015; Rahmanifard
83 and Plaksina, 2019; Sabah et al., 2019; Sun and Ertekin, 2020; Enab and Ertekin, 2021; Gouda et al.,
84 2021), Gaussian Process Regression (GPR) (Knowles, 2006; Zhang et al., 2009; Horowitz et al., 2013)
85 methods, and (3) Deep Learning (DL) methods such as Convolutional Neural Network (CNN) (LeCun et
86 al., 1998; Glorot et al., 2011; Hinton et al., 2012; Chu et al., 2020; Kim et al., 2020; Kim et al., 2021).
87 Physics-based approaches can approximate the original reservoir behavior with lower-order equations
88 to reduce the computational cost. However, they have been so far tested on synthetic, box-shaped

89   models only (de Brito and Durlofsky, 2020a; de Brito and Durlofsky, 2020b) and can become
90   unrepresentative in real fields with often complex structures. ML techniques are widely applied within
91   the context of well control optimization (Ahmadi and Bahadori, 2015; Golzari et al., 2015; Chugh et
92   al., 2016; Guo and Reynolds, 2018; Chen et al., 2020; Zhao et al., 2020) and are shown to provide a
93   reasonably accurate, data-driven SM while considering the reservoir simulator as a black box. The
94   accuracy of ML techniques reduces significantly when the control variables become categorical or
95   integer (Junior et al., 2021). This lower accuracy is mainly due to disregarding spatial features (e.g.
96   well location, type, and trajectory) in a large-scale problem and transforming the inputs to a 1D array
97   (Chu et al., 2020). CNNs methods are an advanced form of ANNs that eliminate the issues associated
98   with the conventional ML techniques by allowing the direct import of multi-dimensional data to the
99   network (LeCun and Bengio, 1995; Behnke, 2003). This allows multi-dimensional input variables such
100   as well locations and types of wells (i.e., injectors and producers) to be directly imported to the
101   network without losing their spatial information providing greater SM prediction performance (Chu
102   et al., 2020; Razak and Jafarpour, 2020). Standard feedforward CNNs are used to predict one, single
103   output while Recurrent Neural Networks (RNNs) such as Long-Short Term Memory (LSTM) can predict
104   time series (Gers et al., 2002; Hua et al., 2019; Sagheer and Kotb, 2019; Liu et al., 2020; Song et al.,
105   2020). CNNs can be applied for both classification and regression problems (Figueiredo, 2003). In this
106   study, standard feedforward CNN is employed as a regression tool to estimate the Net Present Value
107   (NPV) of a reservoir model based on a specified well placement scenario.

108   The employment of surrogate modelling approaches in optimization frameworks has significantly
109   reduced the computational cost of objective function evaluations. However, current surrogate-
110   assisted optimization frameworks provide only a single optimal solution, which lacks flexibility due to
111   often unforeseen operational problems. Hence, the efficient use of surrogate-assisted optimization in
112   providing operational flexibility at a reasonable computational cost remains a critical challenge. This
113   study presents a surrogate-assisted, multi-solution optimization framework to achieve diverse, close-
114   to-optimum well placement scenarios at a reduced computational cost. Following Salehian et al.
115   (2020a), SPSA is used as the optimization algorithm while the diversity in optimal solutions is achieved
116   by multiple, parallel runs of the optimizer with different starting points. CNN is used as the SM, to
117   partly substitute the computationally expensive reservoir model runs during the optimization process
118   by predicting the objective value based on a 2-dimensional map of the location of the vertical wells'
119   tops. An adjusted Latin Hypercube Sampling (aLHS) procedure is developed to generate initial training
120   datasets with diverse well placement scenarios while respecting reservoir boundaries and minimum
121   well spacing constraints. The ensemble learning strategy is used to enhance the accuracy of the SMs
122   prediction as well as to allow estimation of the SM's prediction error for a new data point. The
123   ensemble of CNNs is adaptively updated during the optimization process using newly generated data
124   points with large prediction errors, estimated by the variance of SMs' responses (Cheng and Lu, 2020),
125   to gradually improve the global prediction accuracy of the SMs. The proposed framework has been
126   tested on two representative benchmark case studies (Brugge and Egg field models) while comparing
127   three optimization schemes: 1) "no SM": using full-physics reservoir model only, 2) "offline SM": pre-
128   trained SMs are not updated during the optimization process with no further updates, and 3) "online
129   SM": pre-trained SMs are updated during the optimization process using new data points with
130   maximum prediction errors.

131   This paper is organized as follows: First, problem formulation for well placement optimization using
132   the surrogate-assisted, multi-solution SPSA (SAMS-SPSA) is presented. Next, the framework is tested
133   on the benchmark case studies followed by the discussion of the results and conclusions.

3. Problem Formulation

The objective is to find the optimal values for decision variables (i.e. well locations) that maximize an objective function. Net Present Value (NPV), considering only oil and water production and injection over the presumed life of the reservoir, is the selected objective function, defined as:

$$\underset{\substack{x \in \mathbb{R}^{N_x} \\ m \in \mathbb{R}^{N_m}}}{J}(x,m) = \sum_{n=1}^{S}\left\{\left[\sum_{j=1}^{N_P}(r_o q_{o,j}^n - r_{pw} q_{w,j}^n) - \sum_{k=1}^{N_I}(c_{wi} q_{wi,k}^n)\right] \times \frac{\delta t^n}{(1+b)^{t_n}}\right\} \tag{1}$$

$$subject\ to\ x_i^{min} \le x_i \le x_i^{max}\ ,i = 1,2,\dots,N_x$$

, where x is the N_x dimensional vector of the decision variables; m is the N_m dimensional state vector of the reservoir (e.g. saturation, pressure field); n is the nth time step of the reservoir simulation; S is the total number of simulation steps; δt^n is the length of nth simulation step; t_n is the simulation time at the end of the nth time step; the annual discount rate b is in decimal; and N_P and N_I are the number of producers and injectors, respectively. The cost coefficients r_o, r_{pw}, and c_{wi} are the oil price (USD/STB), the water handling cost (USD/STB), and the water injection cost (USD/STB), respectively. $q_{o,j}^n$ and $q_{w,j}^n$ are the oil and water production rates of well j at time step n in STB/day. $q_{wi,k}^n$ is the water injection rate of well k at time step n in STB/day. x_i^{min} and x_i^{max} are the lower and upper bounds for the ith component of the decision variable vector x. Eq(2) is employed to scale the control variables x from the original domain $[x_{min}, x_{max}]$ to [0, 1] to eliminate the problem of different ranges of decision variables during the optimization process.

$$u_i = \frac{x_i - x_{min,i}}{x_{max,i} - x_{min,i}} \tag{2}$$

Table 1 shows the economic values used for NPV calculation. Simulation runs are conducted using a commercial reservoir simulator (Schlumberger, 2017) to calculate the objective function for the specified set of optimization variables and state vector of the reservoir.

Table 1 - Economic parameters for calculating NPV.

Symbol	Parameter	Value
r_o	Oil Price	50 USD/STB
r_{pw}	Water production cost	6 USD/STB
c_{wi}	Water injection cost	3 USD/STB
b	Discount rate	10% /year

3.1. SPSA Background

Let $J(u_k)$ to be the objective function value at u_k, the N_x dimensional vector of the optimization variables at iteration k. The steepest ascent gradient $g_k(u)$ is defined as the partial derivatives of the objective function $g_k(u) = \frac{\partial J}{\partial u} = \left[\frac{\partial J}{\partial u_1}, \frac{\partial J}{\partial u_2}, \dots, \frac{\partial J}{\partial u_{N_x}}\right]^T$ where $[.]^T$ represents a column vector. SPSA iteratively maximizes the objective function $J(u)$ using the following relationship:

$$u_{k+1} = u_k + \alpha_k \hat{g}_k(u_k) \tag{3}$$

, where $\hat{g}_k(u_k)$ is the stochastically estimated gradient of the objective function and $\alpha_k > 0$ is the step size in the search direction $\hat{g}_k(u_k)$. To calculate $\hat{g}_k(u_k)$, Δ_k is defined as a vector of mutually independent, mean-zero random variables $\left\{\Delta_{k_1}, \Delta_{k_2}, \dots, \Delta_{k_{N_x}}\right\}$ using symmetric ±1 Bernoulli distribution (Spall, 1992), meeting the following conditions:

$$\Delta_{k_i}^{-1} = \Delta_{k_i} \tag{4}$$

$$E\left|\Delta_{k_i}^{-1}\right| = E\left|\Delta_{k_i}\right| = 0 \tag{5}$$

164 , where E represents the expected value. The stochastic gradient $\hat{g}_k(u_k)$ is calculated using Δ_k and a
165 positive scalar c_k:

$$\hat{g}_k(u_k) = \frac{J(u_k + c_k \Delta_k) - J(u_k - c_k \Delta_k)}{2c_k} \times \left[\frac{1}{\Delta_{k_1}}, \frac{1}{\Delta_{k_2}}, \dots, \frac{1}{\Delta_{k_{N_x}}} \right]^T \tag{6}$$

166 The convergence of the SPSA depends on the tuning parameters α_k and c_k, which are particularly
167 important when the objective function is computationally demanding. Spall (1998) suggested the
168 following decaying sequences to calculate α_k and c_k to ensure a gradually refining search:

$$\alpha_k = \frac{a}{(\mathbb{A} + k + 1)^{\vartheta}} \tag{7}$$

$$c_k = \frac{c}{(k + 1)^{\gamma}} \tag{8}$$

169 , where a, c, $\mathbb{A}$, ϑ, and γ are positive real numbers. The values of ϑ and γ are recommended to be
170 0.602 and 0.101 (Spall, 1992). Following Jesmani et al. (2020), $\mathbb{A} = 100$ is used to achieve a more
171 refined search in order to enhance the convergence of the algorithm in the well placement
172 optimization problem with discrete decision variables. Haghighat Sefat et al. (2016) recommended
173 setting $0.1 \leq \alpha_0 \leq 0.5$ and c_{min} (i.e. when $k = k_{max}$) between 0.025 and 0.1 based on the
174 complexity/noise of the search space. Initial sensitivity analysis in this work showed that $\alpha_0 = 0.5$ and
175 $c_{min} = 0.08$ yield a faster convergence and more stable search process. In this work, parallel,
176 independent optimization runs are performed using SPSA algorithm with N_s different starting points
177 in order to achieve N_s diverse solutions with close to optimal objective values.

178 **3.2. Surrogate Modeling using CNN**
179 In this section, the mathematical configuration of Artificial Neural Networks (ANNs, i.e. neural
180 networks with one hidden layer) and deep neural networks (DNNs, i.e networks with multiple layers
181 between inputs and outputs) are discussed. In ANNs, the relationship between input and output is
182 approximated using three layers of neurons: one input layer, one hidden layer, and one output layer.
183 Each of these layers is composed of nodes, also called neurons, which are connected to the neurons
184 of other layers by connections with specified weights. A layer is considered fully connected if all its
185 neurons are interconnected with all the neurons in the adjacent layers (see example in Figure 1).
186 Mathematically, matrix multiplications between 1D input data and weights are performed, and the
187 result is forwarded to the next layer to be used as input. The data flow at a certain layer can be
188 expressed as follows:

$$Y = \sigma(XW + B) \tag{9}$$

189 , where X is the result from the previous layer; Y is the output, which becomes the input X of the next
190 layer; σ is the activation function; W is the weight; and B is the bias, which is used to shift the
191 activation function by a constant value (Rumelhart et al., 1985). Note that in ANNs, both W and B are
192 1D learnable arrays. Activation function σ allows to capture the non-linear relationships between
193 inputs and outputs (Specht, 1990). Table 2 shows four activation functions compared by Chu et al.
194 (2020): the sigmoid and hyperbolic tangent functions have been commonly used in ANNs, while the
195 ReLU function has been mainly used in DNNs (Glorot et al., 2011; Maas et al., 2013).

196 A CNN is a class of DNN, which is commonly used for multi-dimensional regression and image
197 classification/recognition (Goodfellow et al., 2016). CNN is composed of two stages: (1) a convolution
198 stage and (2) a full connection stage. The convolution stage itself consists of multiple, consecutive
199 convolutional layers and pooling layers, where features of the input array are extracted. Batch
200 normalization is performed after each convolutional layer to re-center and re-scale the data to
201 accelerate the training by regularizing the model (Ioffe and Szegedy, 2015). The activation function
202 ReLU is then applied to the convolution outputs of each layer, in the same way as shown in Eq(9).

203   Next, a pooling layer is used to reduce the dimension of the feature map by taking a representative
204   value of each independent sub-cluster covered by the filter, in order to reduce the number of
205   learnable parameters and avoid overfitting (Scherer et al., 2010). Two typical pooling methods exist
206   (1) Max pooling: selecting the maximum value of each sub-cluster as the representative value and (2)
207   Average pooling: calculating the mean value of the elements in each sub-cluster. Following the
208   recommendation by LeCun et al. (2015), average pooling is used in this work. The whole convolution
209   stage f is defined by

$$Y = f(X, \Theta) \tag{10}$$

210   , where $\Theta = [W_1, \dots, W_{N_{cv}}, B_1, \dots, B_{N_{cv}}]$ is the augmented vector of convolution, learnable parameters
211   including weights W (also referred to as a filter) and bias B; N_{cv} is the number of convolutional layers
212   in the convolution stage (subscript cv refers to convolution); and Y is the convolution output called
213   feature map, dimension of which depends on that of input X and the filter size. At the end of the
214   convolution stage, the extracted multi-dimensional feature maps are flattened into a 1D array before
215   being imported to the full connection stage, which consists of a fully connected layer similar to an
216   ANN. Figure 2 shows a schematic summary of a CNN.

217   In this study, CNN is used to predict the NPV based on the imported 2D map of well configurations, in
218   which each grid block with an injector, a producer, or no wells is represented by -1, 1, and 0,
219   respectively. The set of all weights and biases (i.e. W and B, respectively) in all layers of the network
220   are optimized using the backpropagation algorithm (LeCun et al., 1989) to minimize the error function
221   L between the network output (H) and the true outputs (J, i.e. the objective value calculated using
222   the reservoir simulator) using the training data sets. Error function L is defined as

$$L = \|H - J\|_2 \tag{11}$$

223   , where $\|.\|$ represents the l_2-norm. The trained network is used during the optimization process to
224   substitute the actual, time-consuming, reservoir simulation runs.

225   ***Ensemble learning strategy:*** Following previous works (Zerpa et al., 2005; Goel et al., 2007; Zhao and
226   Xue, 2011; Viana et al., 2013; Cheng and Lu, 2020), an ensemble of SMs is trained using different
227   random seeds and the final output is the average of the individual SMs outputs. This approach is
228   expected to enhance the robustness of the SM by alleviating the impact of random seeds on the
229   training performance as well as allowing estimation of the SM's prediction quality for new data points.
230   Note that all SMs are trained using the same training dataset. The mean value of the predicted
231   objective function over an ensemble of SMs (i.e. $\bar{H}$ according to Eq(11)) is maximized during the
232   optimization process, as follows:

$$\max_{x \in \mathbb{R}^{N_x}} J(x) \approx \bar{H}(x) = \frac{1}{N_e} \sum_{k=1}^{N_e} H_k(x) \tag{12}$$

233   , where $H_k(x)$ is the output of the kth SM based on input x (i.e. an iteration of the decision variables)
234   and N_e is the number of SMs.
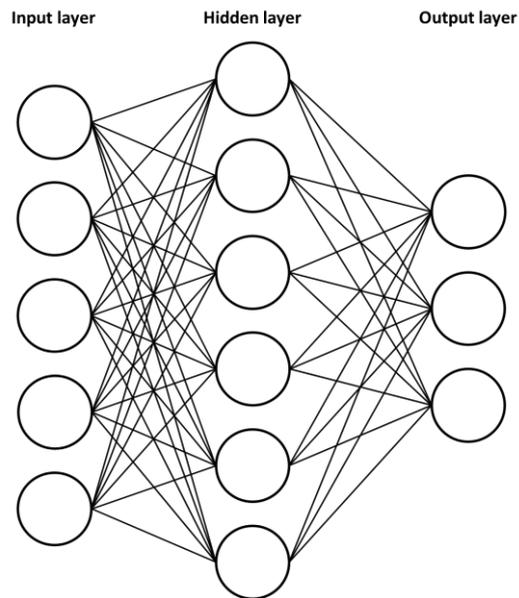
235

236 Figure 1 – Schematic example of a fully connected ANN with 5, 6, and 3 nodes in the input, hidden,
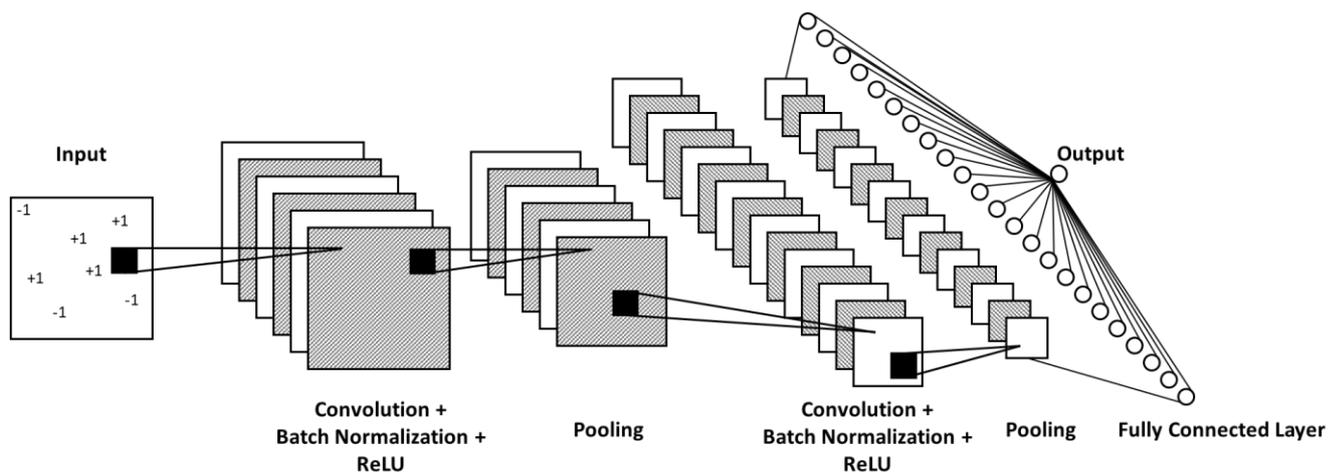237 and output layer, respectively.

238

239 Table 2 – Typical activation functions for neural networks.

Activation Function	Formulation
Sigmoid	$\sigma(x) = \dfrac{1}{1 + e^{-x}}$
Hyperbolic Tangent	$\sigma(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$
Linear	$\sigma(x) = x$
ReLU	$\sigma(x) = \max(0, x)$

240

241

242

243 Figure 2 – Schematic example of the architecture of a CNN.

3.3. Design of Experiments using Adjusted Latin Hypercube Sampling (aLHS)

The initial training of SMs starts with the design of experiments. LHS (McKay et al., 2000) is commonly used due to its space-filling features to generate random training datasets that are evenly distributed over the design space to ensure good coverage. In this study the following constraints should be imposed during both the training datasets generation using LHS, as well as the SPSA iterations:

1. *The wells are located inside the reservoir boundaries:* Following Salehian et al. (2020a), a binary matrix with 0 and 1 elements – representing null and active reservoir grids, respectively – is generated. The well is moved to the nearest active grid if it appears outside the reservoir boundaries (see Algorithm 1 in Appendix A).
2. *A minimum inter-well distance* (see Algorithm 2 in Appendix A): Pair-wise distances are calculated for all wells. If any pair of wells violate the minimum distance limit (R_{min} in Algorithm 2), a fast optimization problem finds the closest acceptable locations for those particular wells while respecting the minimum inter-well distance with all other wells and the reservoir boundaries constraint. In this study, the minimum inter-well distance constraint is set to 2 grid blocks.

The aLHS is composed of the classic LHS followed by the adjustment steps (i.e. Algorithm 1 and 2) to respect the abovementioned constraints. The adjusted training data points are then evaluated using the reservoir simulator to be used for the initial training of the ensemble of SMs. Figure 3 shows the flow diagram of the initial training process. Note that aLHS is also used to generate N_s different starting points for the SPSA as well as for adjusting its solutions during the optimization iterations.
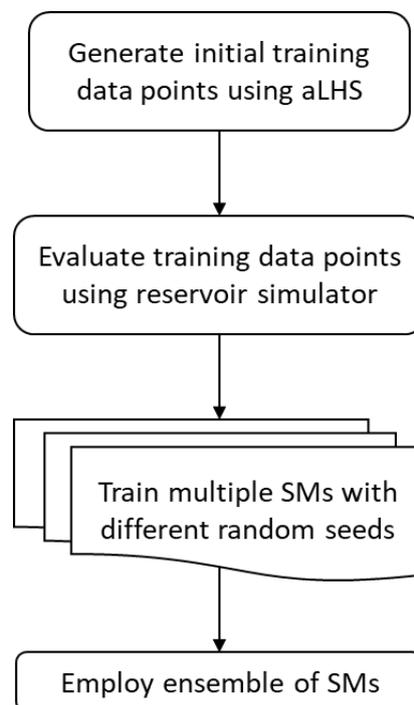
Figure 3 – Initial training of ensemble of SMs.

3.4. SM Quality Assessment

In an offline surrogate modeling scheme, the SM is only trained using the initial training dataset and will be used during the subsequent optimization with no further updates. However, this approach generally leads to a sub-optimal solution due to the SM prediction error (Jin, 2005; Bouzarkouna et

271 al., 2012). The use of an online (also called "adaptive") surrogate modeling scheme is recommended
272 by several studies (Jin, 2011; Razavi et al., 2012; Golzari et al., 2015; Sayyafzadeh, 2017), to iteratively
273 improve SM's prediction accuracy using new data points generated during the optimization process.
274 Following Cheng and Lu (2020), the variance of the ensemble of trained SMs' outputs ($\sigma^2\big(H(.)\big)$)
275 provides an indication of the prediction accuracy for a new data point. In this study, during
276 optimization process and prior to objective function evaluation for each new data point (x_k) the SMs
277 prediction accuracy ($\sigma^2\big(H(x_k)\big)$) is calculated and compared with a threshold value (σ^2_{min}). If
278 $\sigma^2\big(H(x_k)\big) > \sigma^2_{min}$ then the data point (x_k) is evaluated using the reservoir simulator and is added as
279 a new data point to the training dataset, otherwise, the SMs are used for function evaluation. This
280 approach aims to improve the SMs prediction performance in the search space regions with the
281 maximum error (i.e. the least explored regions by the prior reservoir simulation runs) while ensuring
282 accurate and fast objective function evaluation during the optimization process. In this study, σ^2_{min} of
283 0.02 was selected, which is the average of variances of SMs' outputs for the initial training dataset.

4. Surrogate-Assisted, Multi-Solution framework based on SPSA (SAMS-SPSA):

285 Figure 4 shows the summary of the developed SAMS-SPSA framework. First, a set of initial training
286 points are generated using aLHS method and the corresponding outputs are calculated using the
287 reservoir simulator. This dataset is employed to pre-train the ensemble of CNNs with different random
288 seeds. Then, N_s starting points are generated using aLHS followed by N_s parallel runs of the optimizer,
289 to maximize the mean of the ensemble of CNNs' outputs. If the prediction error of the ensemble of
290 CNNs (i.e. the variance of ensemble of trained CNNs' predictions) for a function evaluation is greater
291 than σ^2_{min}, that point is evaluated using the reservoir simulator, added as a new training data point
292 followed by re-training the ensemble of CNNs using the updated dataset. The SAMS-SPSA framework
293 is terminated when the maximum number of optimization iterations is reached (as will be shown later,
294 this number is by far sufficient to converge to optimal solutions). In this study, three optimization
295 schemes are compared:

296 - S1: No SM: classic approach, using full-physics reservoir model only.
297 - S2: Offline SM: optimization is performed using an ensemble of pre-trained SMs with no
298   further updates.
299 - S3: Online SM (proposed approach): pre-trained SMs are updated during the optimization
300   process when their prediction error is high for a new data point.

301 All schemes are compared when performing three multi-start, parallel optimization runs (i.e. $N_s = 3$)
302 from identical starting points. For the SM-assisted schemes (S2 and S3) in the Brugge and Egg field
303 case studies, respectively, 1000 and 700 initial training data points are generated to pre-train an
304 ensemble of 10 CNNs with different random seeds (i.e. $N_e = 10$ in Eq.(12)). Note that N_s and N_e are
305 user-defined, empirical parameters, and the choice of these parameters did not seem to strongly
306 impact the performance of framework, though further research to formalize this strategy will be
307 useful. The hyperparameters and architectures for the employed CNNs are shown in Tables B1, B2,
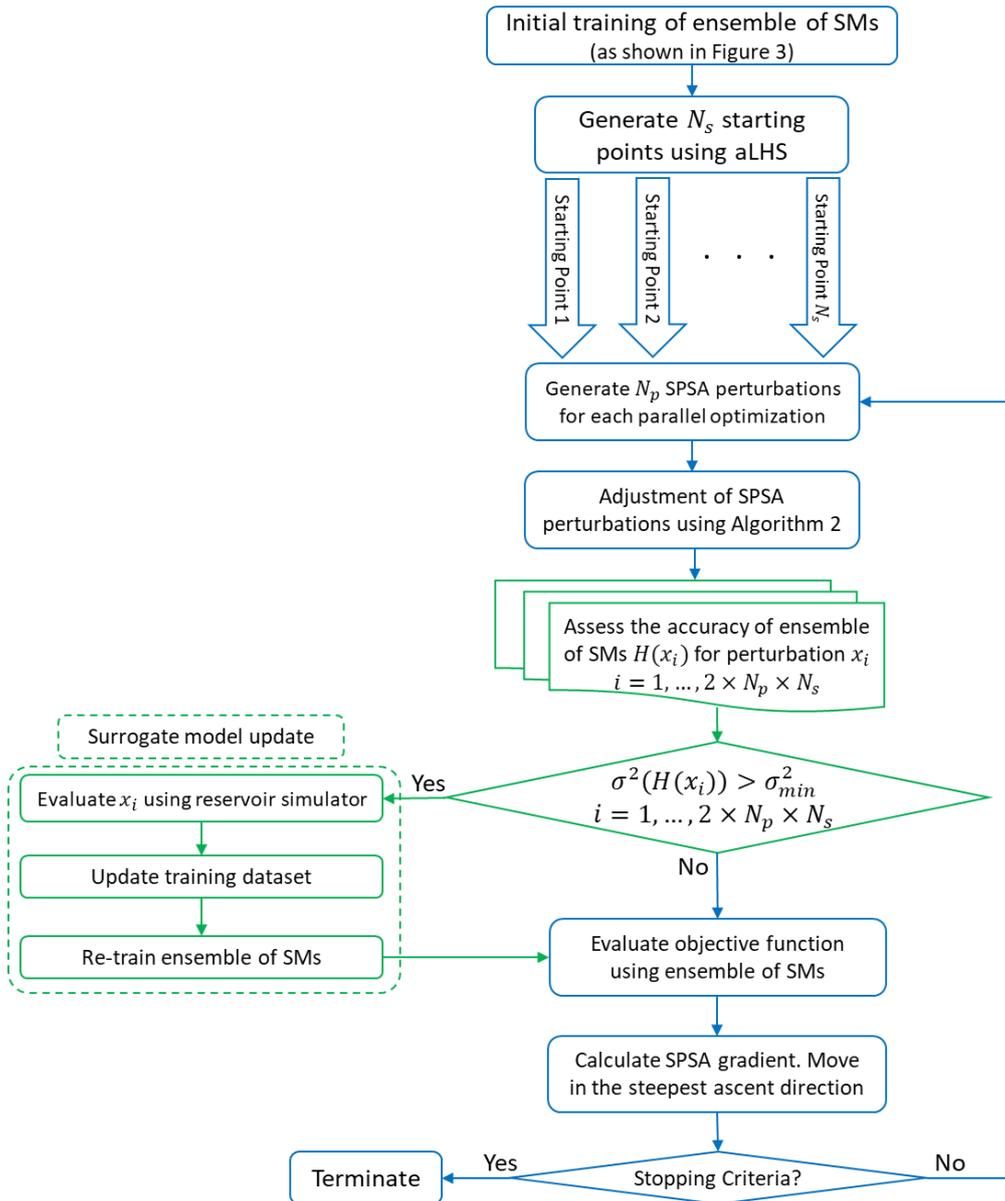308 and B3 in Appendix B.

309

310

Figure 4 – Flow diagram of the SAMS-SPSA framework.

5. Case study 1 – Brugge model

The Brugge field model is a benchmark reservoir model, consisting of $139 \times 48 \times 9$ grid blocks with a relatively heterogeneous permeability distribution (Chen et al., 2010). The original model consists of 20 producers and 10 injectors. Five vertical producers and five vertical injectors are kept from the original model in this work, due to the limited computational resources. The wells are completed in all nine reservoir layers. The total production time is set to 30 years. The producers are each operated by a fixed bottom-hole pressure of $725\ psi$, while the injectors are each operated by a fixed water injection rate of $6289\ STB/day$. The producers are shut when their water cut exceeds 90% since they stop being profitable according to the economic parameters listed in Table 1. Figure 5 shows the top structure of the model with an initial set of well locations. A single realization (the most likely scenario corresponding to $P50$ recovery based on the initial well locations) of the Brugge model is considered for this initial testing of the proposed framework. More information on the reservoir rock and fluid properties of the Brugge model can be found in Peters et al. (2010). The top (i, j) location coordinates of the wells (i.e. $10 \times 2 = 20$ decision variables) are optimized using 100 iterations of the optimization

325   algorithm. Note that all wells start operating at time zero, which means that no drilling sequence is
326   assumed during the optimization process.
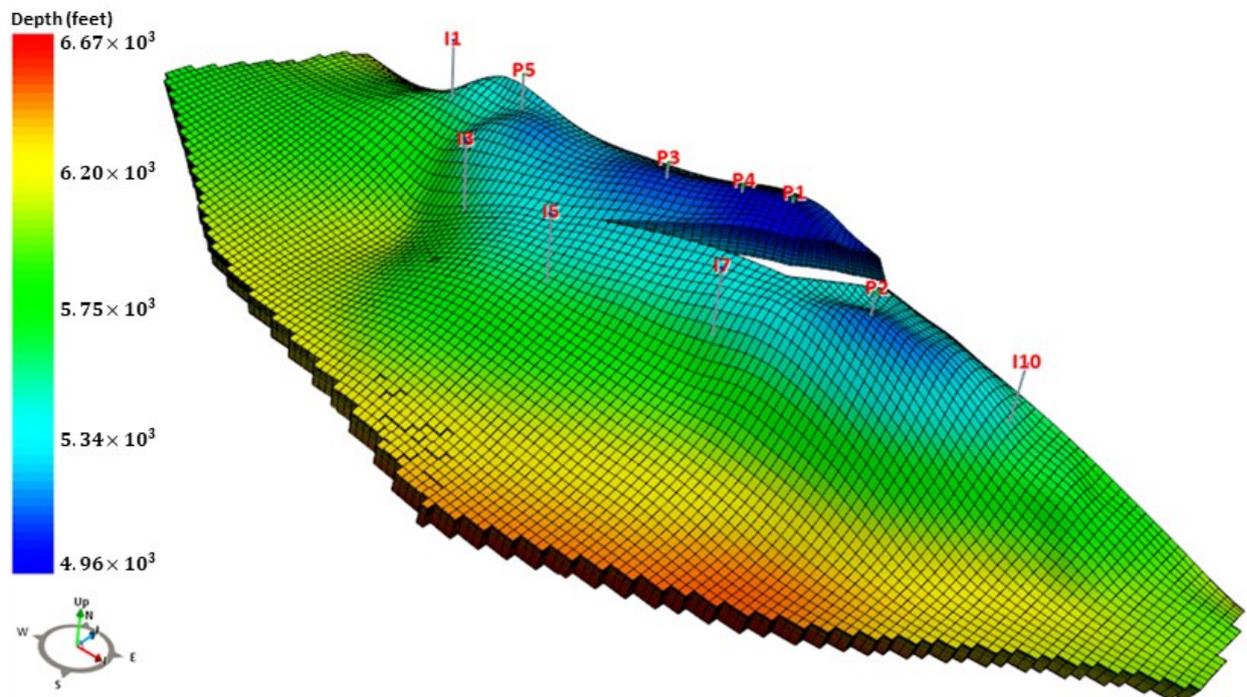
327

328   Figure 5 – Top structure of the Brugge model.

329   ***Brugge-S1:*** Figure 6 shows the improvement of the objective value of three parallel optimization runs
330   during well placement optimization using scheme S1. The oscillations in the NPV are due to the
331   minimum inter-well distance constraint imposed by the adjustment method (see Algorithm 2 in
332   Appendix A) in the objective function definition. The NPV of the starting points (i.e. L_1, L_2, and L_3)
333   respectively, with non-optimal well locations is 1.66×10^9, 1.70×10^9, and $1.81 \times 10^9 \; USD$, which
334   was improved to 2.214×10^9, 2.258×10^9, and $2.232 \times 10^9 \; USD$ after well placement optimization
335   with total of 2100 reservoir simulation runs (note that the NPV of the base case with initial well
336   locations shown in Figure 5 is $1.66 \times 10^9 \; USD$). Figure 7 shows the representative well placement
337   solutions, named S_1L_1, S_1L_2, and S_1L_3, where S_1 denotes the S1 optimization scheme. It is worth
338   noting that an optimization run with a lower initial objective value can potentially provide better final
339   performance (e.g. $J(L_2) < J(L_3)$, but $J(S_1L_2) > J(S_1L_3)$), showing the advantage of multi-start
340   optimization runs in a more efficient exploration of the search space.

341
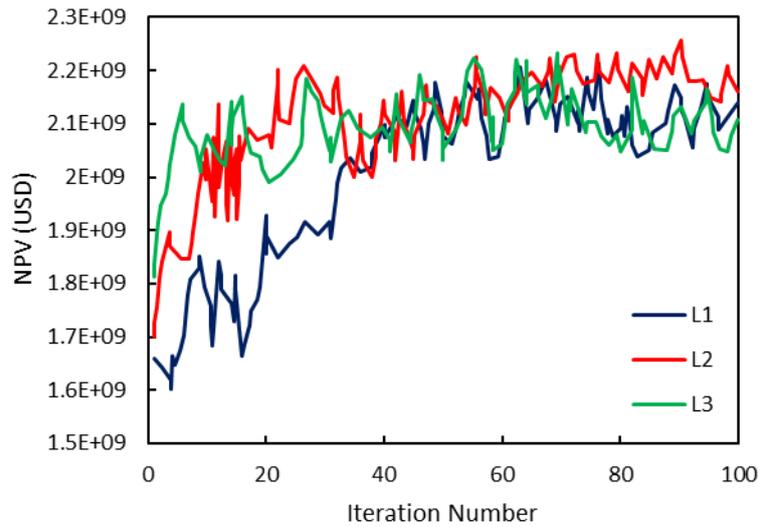
342

343 Figure 6 - Objective value of three parallel runs during well placement optimization of Brugge model-
344 using scheme S1.
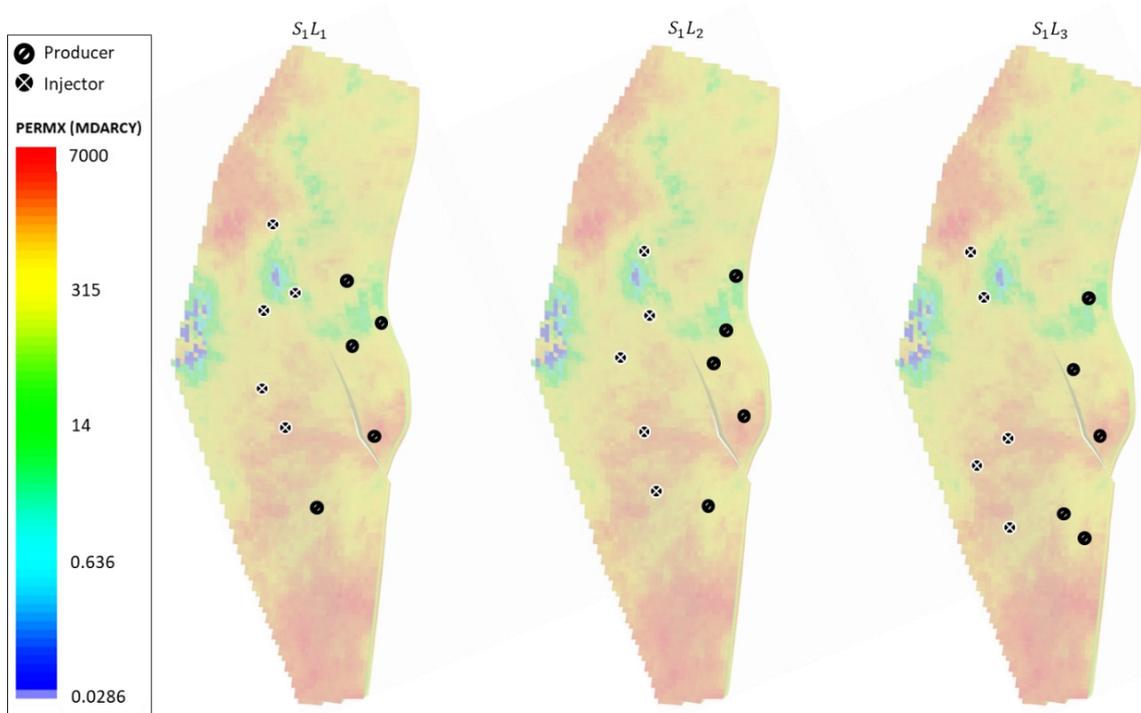
345

346 Figure 7 - Three optimal well placement solutions for the Brugge model obtained using the S1
347 scheme.

348 **Brugge-S2:** First, 70% and 30% of the initial training dataset (of 1000 points), respectively, are used to
349 train and validate a single CNN to test the sufficiency of the training data as well as the suitability of
350 the training process. Note that the training cost of CNN (around 30 seconds) is negligible as compared
351 to the simulation time, which is in the range of 10 minutes per run. The high R^2 between the predicted
352 and the true values (Figure 8) and the declining loss function during the validation process (Figure 9)
353 indicate that the SM is adequately trained with sufficient data and no overfitting. This assessment was
354 repeated with two other random seeds, as shown in Table 3. The formulations for R^2 and $RMSE$ are
355 provided in Appendix C.

356 After ensuring the adequacy of the training dataset and process, 100% of the training dataset is used
357 to train an ensemble of CNNs (see Figure 3). Subsequently, 100 iterations of the well placement
358 optimization are performed in the S2 scheme with no further SM updates. Figure 10 shows the optimal
359 well placement solutions obtained by three multi-start optimization runs using the S2 scheme.
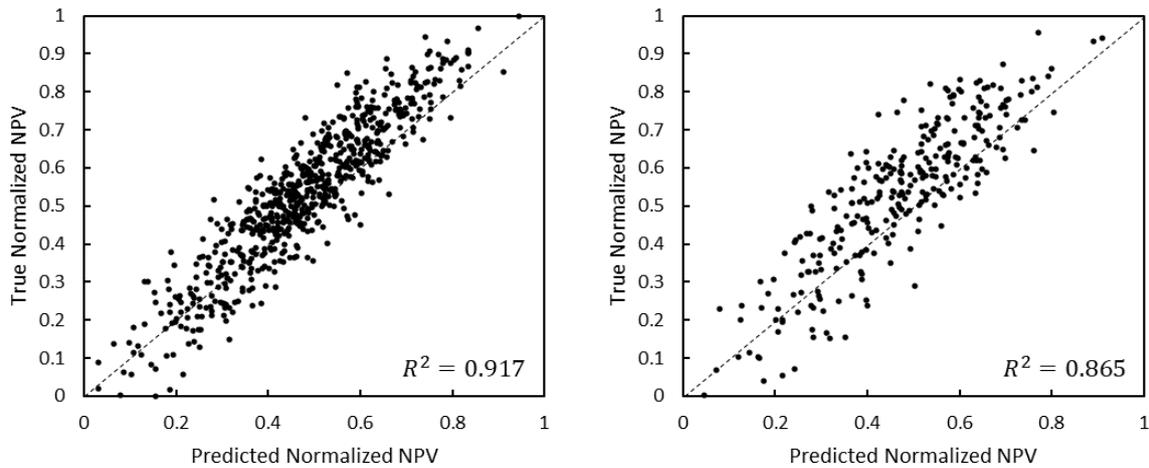
360

361 Figure 8 – Training (left) and validation (right) performance of a single CNN using 700 and 300 data
362 points, respectively (Brugge model example).
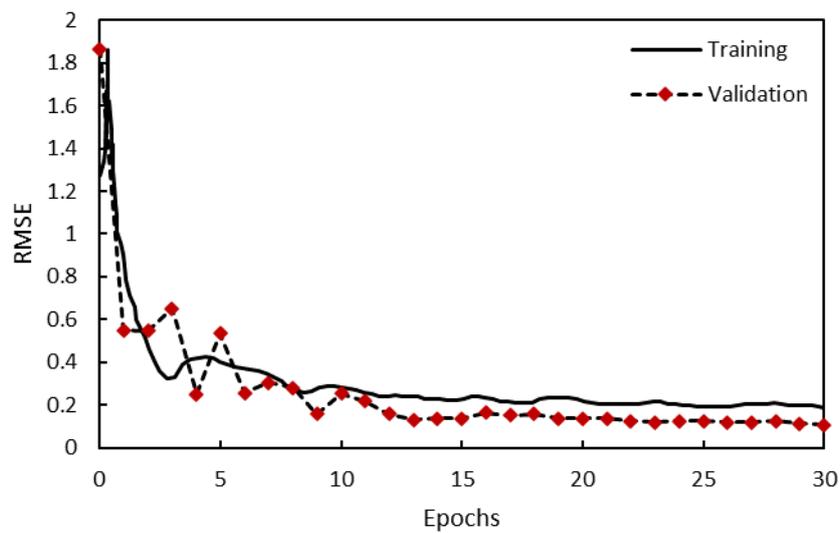
363

364 Figure 9 – Loss function with training epochs for a single CNN (Brugge model example).
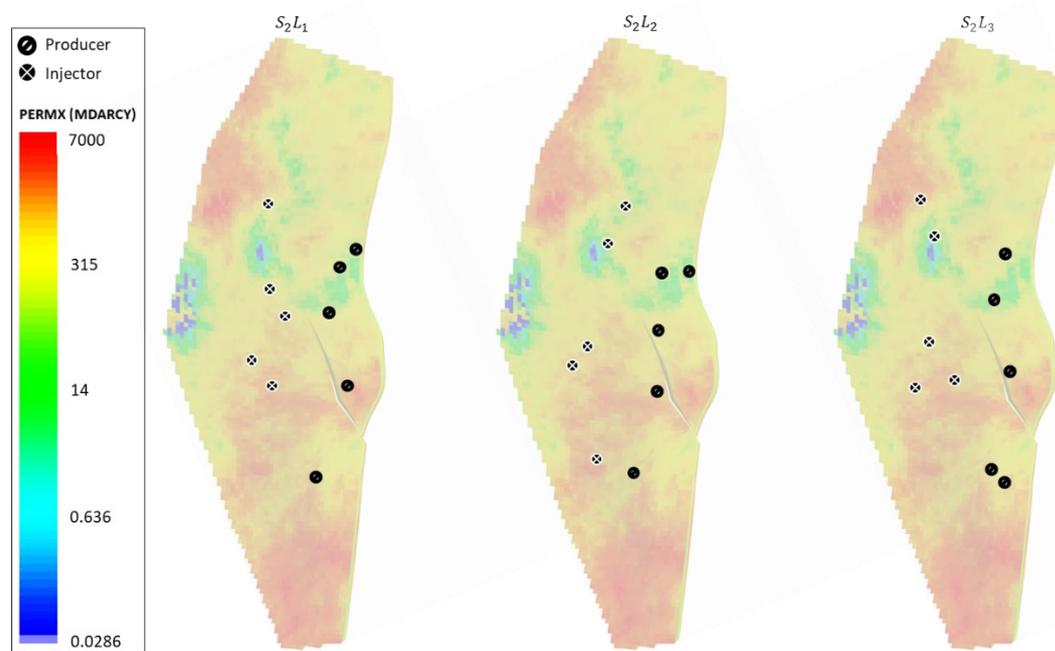
365

366   Table 3 – Pre-assessment performance of a single CNN (Brugge model example).

| Random seed	Parameter	Value
Seed 1		
	R^2 Training	0.9167
	R^2 Validation	0.8648
	RMSE Validation	0.1265
Seed 2		
	R^2 Training	0.9178
	R^2 Validation	0.9076
	RMSE Validation	0.1116

367

368

369   Figure 10 - Three optimal well placement solutions for the Brugge model obtained by the S2 scheme.

370

371   ***Brugge-S3:*** Figure 11 shows the objective value of the three multi-start optimization runs using
372 scheme S3, where the ensemble of pre-trained SMs (same as scheme S2) was updated 42 times (i.e.
373 42 function evaluations, out of the total number of 1800 objective function evaluations, was
374 performed using reservoir simulator) during the optimization process, resulting in 1042 total
375 simulation runs (including 1000 runs for evaluating initial training dataset). Figure 12 shows three
376 optimal well placement solutions obtained by scheme S3. A very similar final objective value is
377 observed by multiple optimal solutions while all schemes offer a reasonable degree of flexibility in the
378 well locations (Figure 7, Figure 10, and Figure 12). Table 4 compares the initial and final objective
379 values of the solutions obtained as well as the number of full simulation runs required at each scheme.
380 The SM-assisted optimization schemes (S2 and S3) reduced the number of simulation runs by 52% and
381 50%, respectively. Very close-to-optimum objective values were achieved by S3 where the average
382 NPV is only 0.54% lower than S1. The prediction accuracy of the offline SM reduces significantly for
383 new data points which resulted in the S2 scheme converging to sub-optimal solutions with an average
384 NPV of 3.85% lower than S1, even when an ensemble of SMs is used.
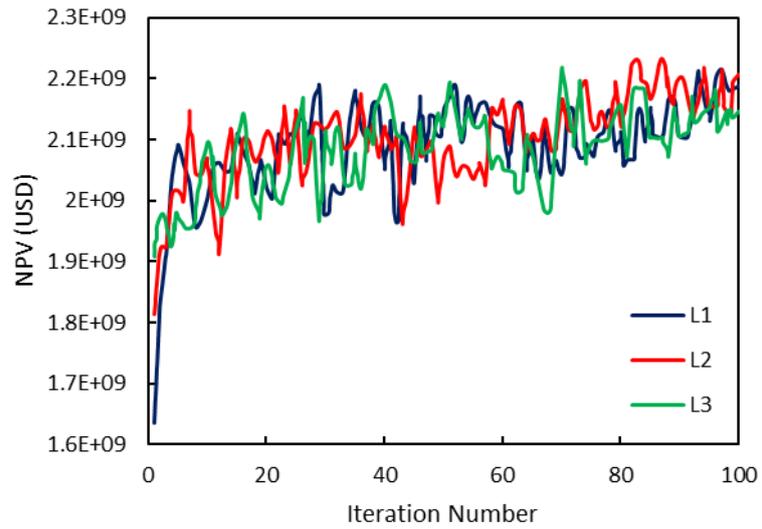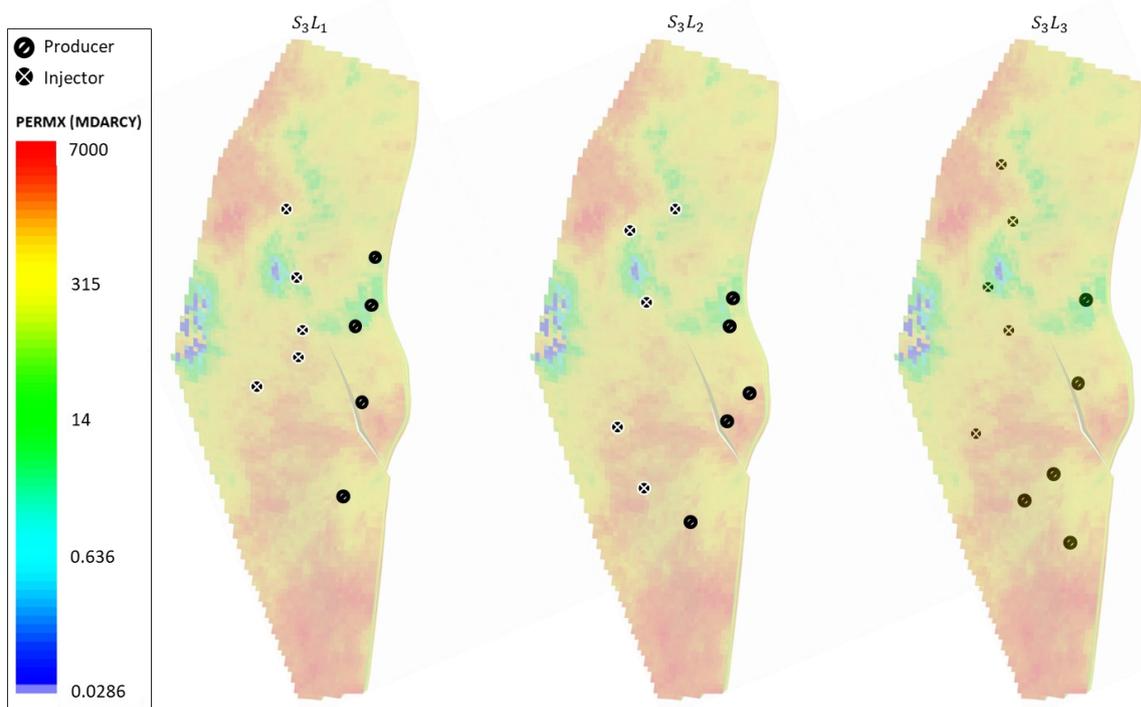
385

386   Figure 11 - Objective value of three parallel runs during well placement optimization of Brugge
387   model; scheme S3.

388

389

390   Figure 12 - Three optimal well placement solutions for the Brugge model obtained by the S3 scheme.

Table 4– Summary of the initial and final objective values with the number of simulations using each surrogate modeling scenario in the Brugge model.

|	Initial NPV (USD)	Final NPV (USD)	Number of simulations
S1: Objective function evaluation using reservoir simulator (no SM)			
L_1	1.66×10^9	2.214×10^9	2100
L_2	1.70×10^9	2.258×10^9	
L_3	1.81×10^9	2.232×10^9	
S2: Objective function evaluation using an ensemble of offline CNNs			
L_1	1.61×10^9	2.153×10^9	1000
L_2	1.63×10^9	2.149×10^9	
L_3	1.68×10^9	2.142×10^9	
S3: Objective function evaluation using an ensemble of online CNNs			
L_1	1.63×10^9	2.220×10^9	1042
L_2	1.81×10^9	2.232×10^9	
L_3	1.91×10^9	2.216×10^9	

6. Case Study 2 – Egg model

The Egg reservoir model is a publicly available, 3D channelized benchmark case study consisting of 60 × 60 × 7 grid blocks, of which 18553 are active. The model contains 8 injectors and 4 producers. Figure 13 shows the horizontal permeability of the top layer for a single geological realization of the field and an initial set of well locations. Similar to the previous case study, a single realization corresponding to $P50$ recovery, based on the initial well locations, is considered. Detailed information on the reservoir rock and fluid properties of the Egg model can be found in Jansen et al. (2014). The field production period is set to 10 years. The producers each operate at the constant BHP of 5727 psi, and they are shut when their water cut reaches 90%. The injectors each operate by their fixed water injection rate of 500 STB/day. The top (i, j) locations of 12 vertical wells are optimized during 100 iterations of the SPSA algorithm resulting in 12×2=24 decision variables while assuming no drilling sequences.
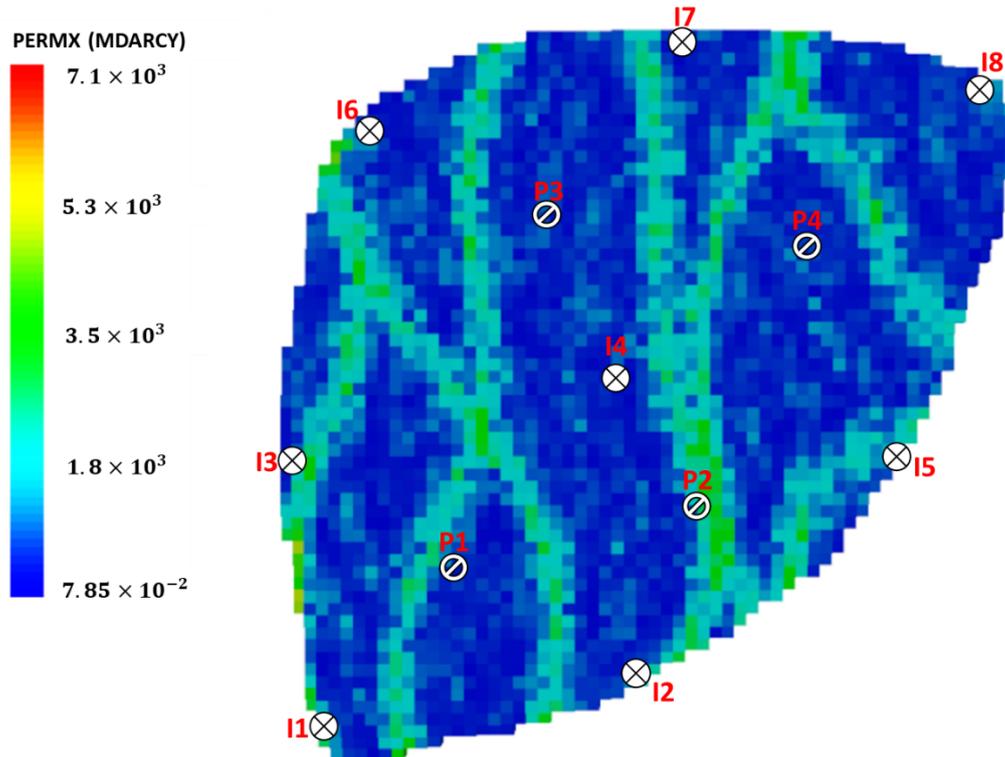
Figure 13 - Permeability distribution of the top layer and the base case well locations of the Egg model.

Egg-S1: The scheme S1 was performed using 2100 simulation runs to improve the NPV of the starting three points L_1, L_2, and L_3, respectively, from 1.55×10^7, 1.43×10^7, and 1.54×10^7 USD, to 1.76×10^7, 1.81×10^7, and 1.80×10^7 USD, as shown in Figure 14 (note that the NPV of the base case with initial well locations shown in Figure 13 is 1.54×10^7 USD). Figure 15 shows the optimal well placement solutions (named S_1L_1, S_1L_2, and S_1L_3) obtained by scheme S1.
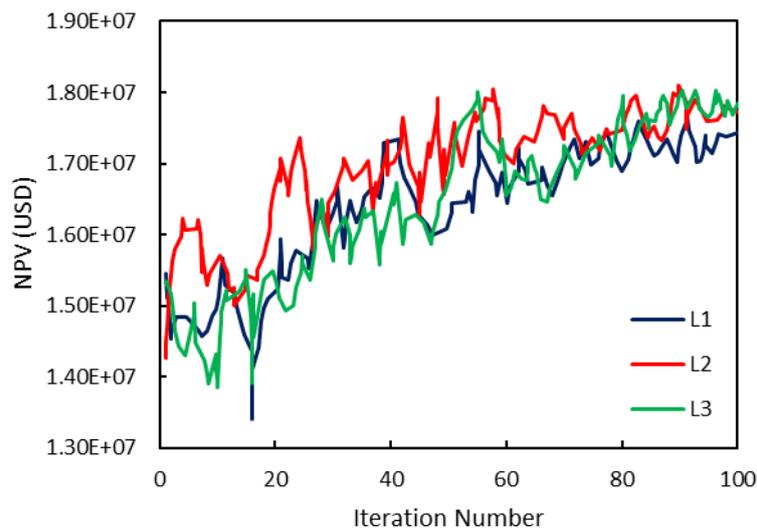
Figure 14 - Objective value of three parallel runs during well placement optimization of Egg model using scheme S1.

416
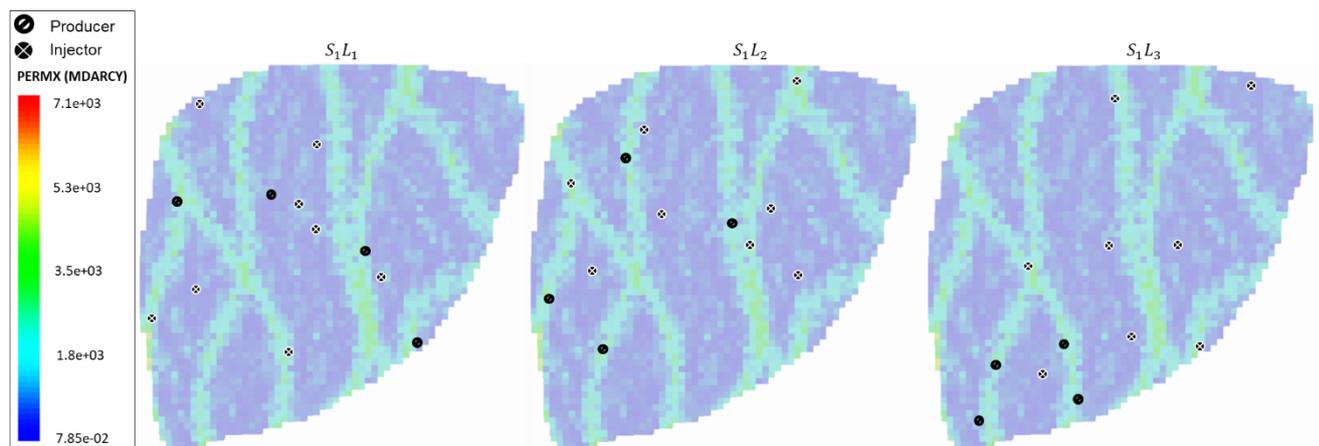
418   Figure 15 - Three optimal well placement solutions for the Egg model obtained by scheme S1.

419   ***Egg-S2:*** Similar to the previous case study, a single SM pre-assessment was first performed by taking
420   500 and 200 data points (of the 700 initial training dataset) as the training and validation subset,
421   respectively. The relatively high R^2 in training and validation processes (Figure 16) as well as the non-
422   increasing loss function during the validation process (Figure 17) shows that the SM is sufficiently
423   accurately trained with no overfitting. The test was repeated with a different random seed and the
424   results are summarized in Table 3. The initial training of the ensemble of CNNs (see Figure 3) was then
425   performed using 100% of the training dataset. The ensemble of CNNs was then employed in the well
426   placement optimization using scheme S2. Figure 19 shows the optimal well placement solutions
427   obtained by the S2 scheme.

428
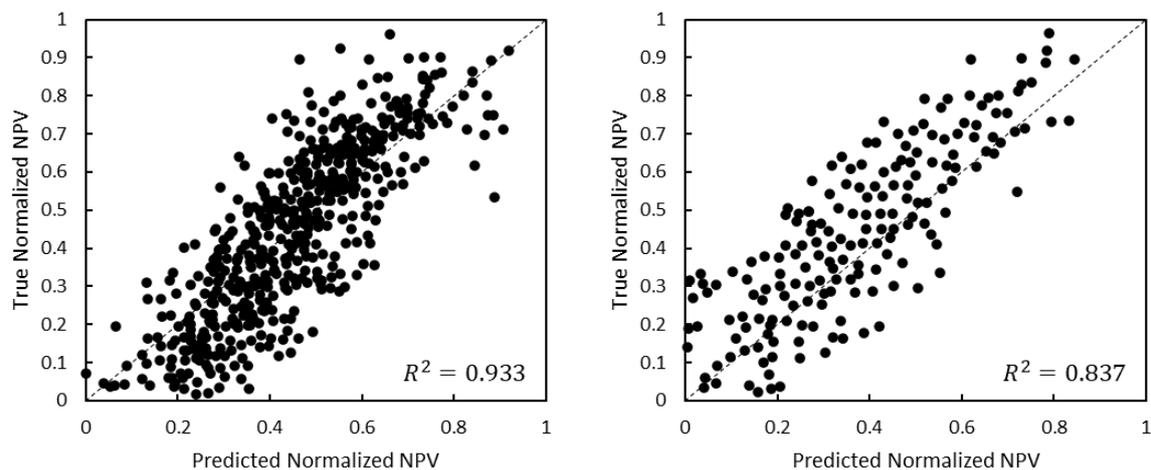
429

430   Figure 16 - Training (left) and validation (right) performance of a single CNN using 500 and 200 data
431   points, respectively (Egg model example).

432
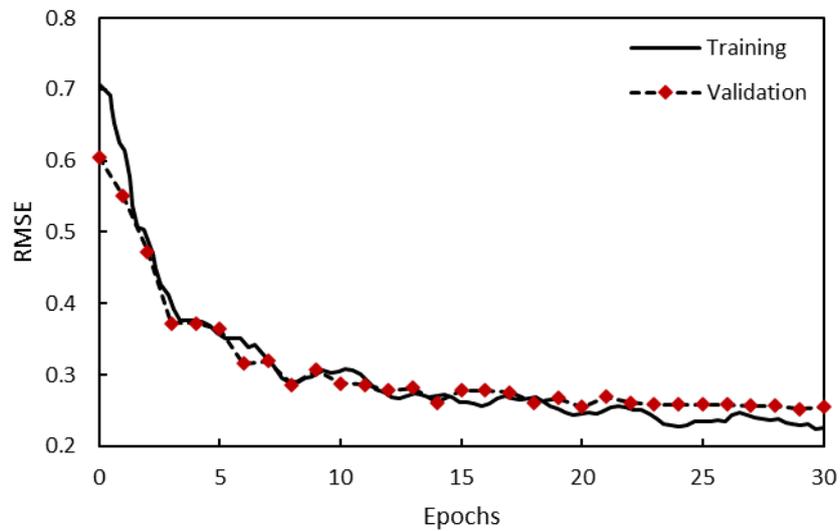
433

434   Figure 17 - Loss function with training epochs for a single CNN (Egg model example).

435

436   Table 5 – Pre-assessment performance of a single CNN (Egg model example).

Random seed	Parameter	Value
	R^2 Training	0.9326
Seed 1	R^2 Validation	0.8372
	RMSE Validation	0.2438
	R^2 Training	0.8206
Seed 2	R^2 Validation	0.7542
	RMSE Validation	0.2817

437
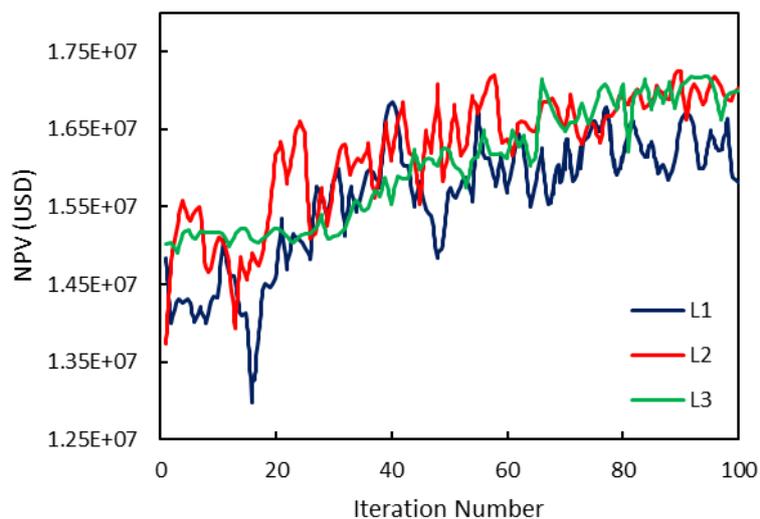
438

439   Figure 18 - Objective value of three parallel runs during well placement optimization of Egg model
440   using scheme S2.
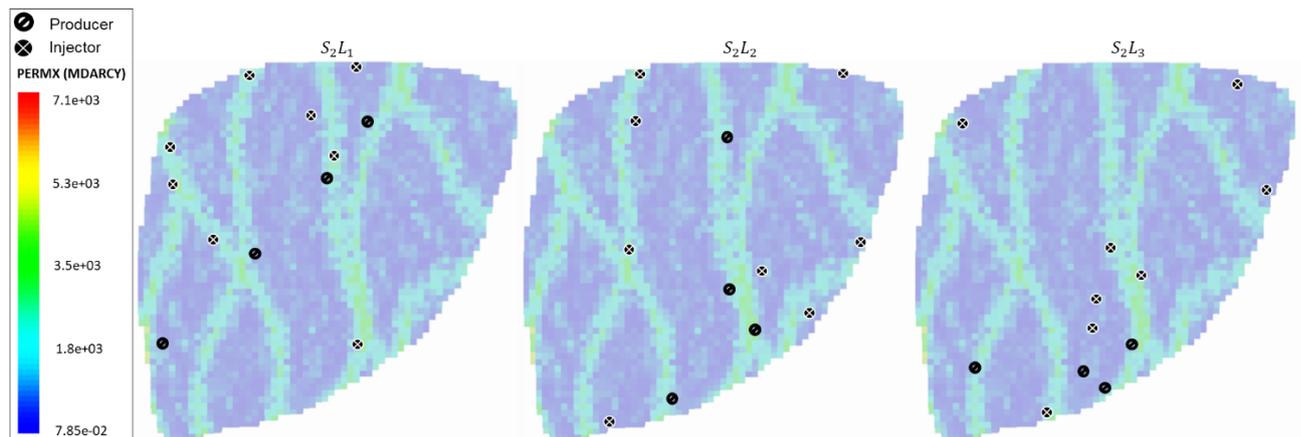
Figure 19 - Three optimal well placement solutions for the Egg model obtained by the scheme S2.

Egg-S3: The ensemble of pre-trained SMs (same as scheme S2) were used in three multi-start optimization runs using scheme S3, as shown in Figure 20. Figure 21 shows three optimal well placement solutions obtained by scheme S3. The ensemble of SMs was updated 154 times during the optimization process, resulting in 854 total simulation runs (including 700 runs used to evaluate the initial training dataset). Different optimal well locations with similar final objective values were obtained by all schemes (Figure 15, Figure 19, and Figure 21). Table 6 shows the initial and final objective values of the solutions obtained as well as the number of simulations required at each scheme. In offline SM, multiple solutions were obtained with 66% less computation time but 4.8% lower average objective values as compared to the no-SM approach, whereas in online SM, closer-to-optimum objective values (3.1% lower compared to S1) were obtained with 59% less computation time. These results are in line with case study 1, showing the advantage of the proposed SAMS-SPSA framework in achieving multiple, diverse optimal solutions with reasonable computational efficiency.

Note that at this stage a single geological realization is considered for a proof-of-concept study and the framework can be extended to perform robust optimization while considering reservoir description uncertainties via importing multi-dimensional inputs (e.g. porosity, permeability maps). SPSA is employed as the optimizer, while the developed framework can also be used with other ensemble-based optimization algorithms. Standard CNNs are used as the SM to partially substitute the computationally expensive reservoir simulation models during the optimization process, though further research to apply time series forecasting methods in optimization process will be useful.
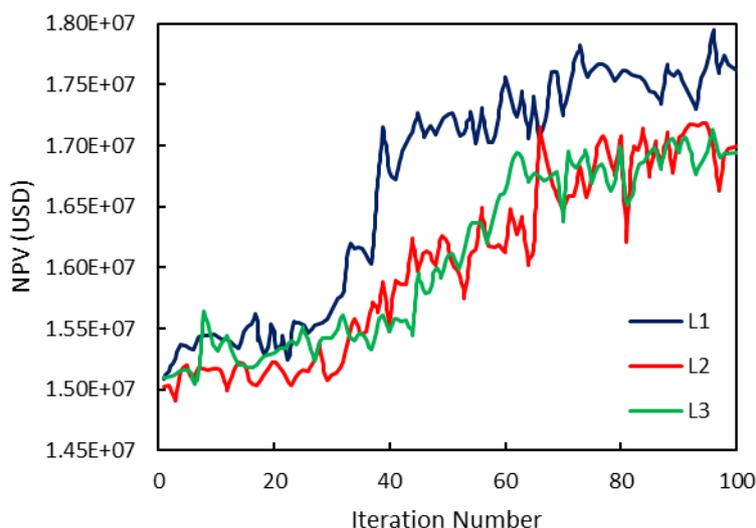
462

463 Figure 20 - Objective value of three parallel runs during well placement optimization of Egg model
464 using scheme S3.

465
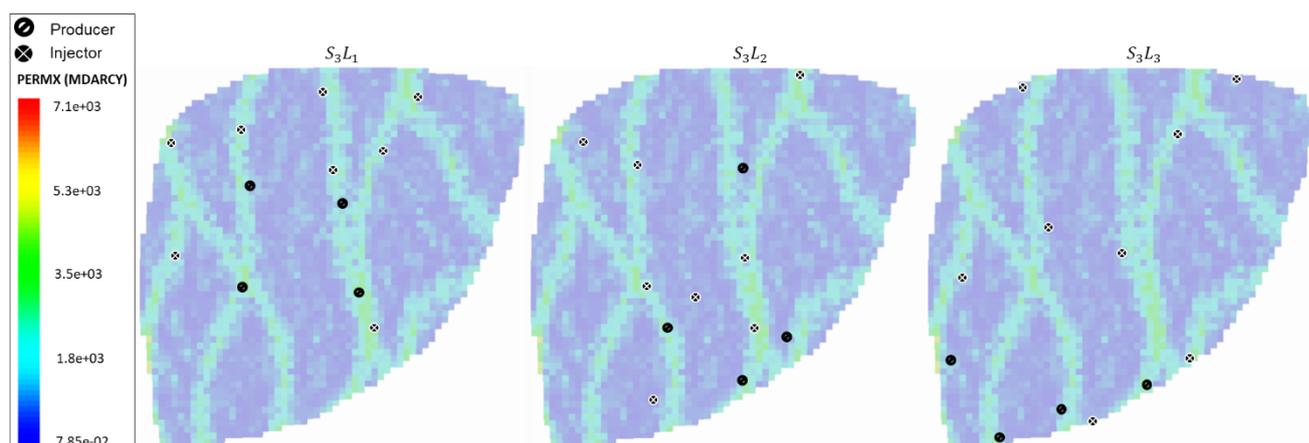
466

467 Figure 21 - Three optimal well placement solutions for the Egg model obtained by the scheme S3.

468

469 Table 6 – Summary of the initial and final objective values with the number of simulations using each
470 surrogate modeling scenario in the Egg model.

	Initial NPV (USD)	Final NPV (USD)	Number of simulations
S1: Objective function evaluation using reservoir simulator (no surrogate model)			
L_1	1.55×10^7	1.763×10^7	2100
L_2	1.43×10^7	1.811×10^7	
L_3	1.54×10^7	1.803×10^7	
S2: Objective function evaluation using an ensemble of offline CNNs			
L_1	1.48×10^7	1.686×10^7	700
L_2	1.37×10^7	1.719×10^7	
L_3	1.50×10^7	1.710×10^7	
S3: Objective function evaluation using an ensemble of online CNNs			

L_1	1.51×10^7	1.792×10^7	854
L_2	1.50×10^7	1.711×10^7	
L_3	1.51×10^7	1.705×10^7	

7. Conclusions

A surrogate-assisted, multi-solution optimization framework is developed to provide operational flexibility by offering multiple, diverse well placement scenarios through multi-start, parallel optimization runs at a reasonable computational cost. The proposed framework was tested on the Brugge and Egg field benchmark case studies and compared with two alternative optimization schemes: no SM (i.e. reservoir simulator only) and offline SM. The following conclusions are drawn:

- Multiple, diverse well placement solutions with close-to-optimum objective values were obtained, demonstrating the advantage of multi-start optimization in a more efficient exploration of the search space.
- CNN, coupled with the ensemble learning strategy, successfully mimicked the reservoir simulator in predicting the objective value using the input well configuration maps.
- Both offline and online surrogate modeling approaches significantly reduced the number of actual reservoir simulation runs required for the well location optimization process. The online SM approach outperformed the offline one due to the continuous improvement in its prediction performance during the optimization process when regions of the search space with minimal prior information were discovered.
- The developed framework provides the much-needed flexibility to field operators by offering them multiple optimal and diverse well placement scenarios at a significantly lower computation cost, as compared to the classical approaches of using reservoir simulator only.

8. Appendix A – Algorithms

Algorithm 1 – Pseudo-code accounting for irregular boundaries during well location optimization

Assume *that binary matrix Ω corresponds to a reservoir top's map, where:*

$$\Omega(i,j) = \begin{cases} 1, if \ (i,j) is \ an \ active \ grid \ of \ the \ reservoir \\ 0, if \ (i,j) is \ a \ grid \ outside \ of \ the \ reservoir \end{cases}$$

N_{well} denotes the number of wells, $\Omega_\alpha(i,j)$ represents an iteration of the location of well α.

Do for *$\alpha = 1, 2, \dots, N_{well}$*
- *If $\Omega_\alpha(i,j) = 1$ then*
  - *Well k is inside the reservoir boundaries.*
- *If $\Omega_\alpha(i,j) = 0$ then*
  - *Find the nearest element of matrix Ω such that $\Omega(i',j') = 1$ and*
    $$d[\Omega(i,j), \Omega(i',j')] = \sqrt{\left(\Omega(i) - \Omega(i')\right)^2 + \left(\Omega(j) - \Omega(j')\right)^2} \ is \ minimum.$$
  - *Replace $\Omega(i,j)$ with $\Omega(i',j')$.*

End do

Algorithm 2 – Pseudo-code for the adjustment procedure for the vector of well locations

---

Assume *a vector of decision variables* $\Psi = [\psi_1(i,j), \psi_2(i,j), \dots, \psi_{N_{well}}(i,j)]$, *where* $\psi_\alpha(i,j)$
represents an iteration of the location of well α.

Do for α *and* β, *where* $\alpha, \beta = 1, 2, \dots, N_{well}$ *and* $\alpha \neq \beta$
    Calculate Euclidean distance between two wells:

$$d[\psi_\alpha(i,j), \psi_\beta(i,j)] = \sqrt{(\psi_\alpha(i) - \psi_\beta(i))^2 + (\psi_\alpha(j) - \psi_\beta(j))^2}$$

If $d[\psi_\alpha(i,j), \psi_\beta(i,j)] < R_{min}$ *then*
    Find $\widehat{\Psi} = [\hat{\psi}_1(i,j), \hat{\psi}_2(i,j), \dots, \hat{\psi}_{N_{well}}(i,j)]$ *such that:*

$$min(\widehat{\Psi} - \Psi)^2$$

    subject to two constraints:
    (1) $\forall \hat{\psi}_\alpha(i,j), \hat{\psi}_\beta(i,j) \in \widehat{\Psi}$ *and* $\alpha \neq \beta$

$$d[\hat{\psi}_\alpha(i,j), \hat{\psi}_\beta(i,j)] \geq R_{min}$$

    (2) $\forall \hat{\psi}_\alpha(i,j) \in \widehat{\Psi}$

$$\Omega_\alpha(i,j) = 1 \text{ (See algorithm 1)}$$

    End if
End do
Replace $\widehat{\Psi}$ *with* Ψ.

494

495   **9. Appendix B – Hyperparameters and Architecture of CNN**

496

497   Table B1 – Hyperparameters used for CNN in the Brugge and Egg model examples.

Hyperparameter	Value
Max epoch	30
Number of iterations per epoch	5
Mini-batch size	128
Initial learning rate	0.003

498

499   Table B2 – CNN architecture in the Brugge model example.

Layer Number	Layer Type	Output Dimension
1	Input Layer	$139 \times 48 \times 1$
2	Convolution Layer	$139 \times 48 \times 8$
3	Batch Normalization	$139 \times 48 \times 8$
4	Activation (ReLU)	$139 \times 48 \times 8$
5	Average Pooling	$69 \times 24 \times 8$
6	Convolution Layer	$69 \times 24 \times 16$
7	Batch Normalization	$69 \times 24 \times 16$
8	Activation (ReLU)	$69 \times 24 \times 16$
9	Average Pooling	$34 \times 12 \times 16$
10	Convolution Layer	$34 \times 12 \times 32$
11	Batch Normalization	$34 \times 12 \times 32$
12	Activation (ReLU)	$34 \times 12 \times 32$
13	Average Pooling	$17 \times 6 \times 32$

14	Convolution Layer	$17 \times 6 \times 64$
15	Batch Normalization	$17 \times 6 \times 64$
16	Activation (ReLU)	$17 \times 6 \times 64$
17	Average Pooling	$8 \times 3 \times 64$
18	Flatten Layer	1×1536
19	Fully Connected Layer	1×1
20	Regression Output	1×1

500

501

502   Table B3 – CNN architecture in the Egg model example.

Layer Number	Layer Type	Output Dimension
1	Input Layer	$60 \times 60 \times 1$
2	Convolution Layer	$60 \times 60 \times 8$
3	Batch Normalization	$60 \times 60 \times 8$
4	Activation (ReLU)	$60 \times 60 \times 8$
5	Average Pooling	$30 \times 30 \times 8$
6	Convolution Layer	$30 \times 30 \times 16$
7	Batch Normalization	$30 \times 30 \times 16$
8	Activation (ReLU)	$30 \times 30 \times 16$
9	Average Pooling	$15 \times 15 \times 16$
10	Convolution Layer	$15 \times 15 \times 32$
11	Batch Normalization	$15 \times 15 \times 32$
12	Activation (ReLU)	$15 \times 15 \times 32$
13	Average Pooling	$7 \times 7 \times 32$
14	Convolution Layer	$7 \times 7 \times 64$
15	Batch Normalization	$7 \times 7 \times 64$
16	Activation (ReLU)	$7 \times 7 \times 64$
17	Average Pooling	$7 \times 7 \times 64$
18	Flatten Layer	1×3136
19	Fully Connected Layer	1×1
20	Regression Output	1×1

503

504 **10. Appendix C – Regression Assessment Metrics**

505 A root mean squared error (RMSE) and R^2, respectively, is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\big(H(x_i) - J(x_i, m)\big)^2}{N}} \tag{C1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\big(H(x_i) - J(x_i, m)\big)^2}{\sum_{i=1}^{N}\big(\bar{J} - J(x_i, m)\big)^2} \tag{C2}$$

506 , where x is the state vector of input data, J is the true output based on the state vector of the reservoir
507 m, $\bar{J}$ is the mean of true outputs, H is the SM response, and N is the number of data points.
508 Acknowledgments

11. References

Ahmadi, M.-A. and Bahadori, A., 2015. A LSSVM approach for determining well placement and conning phenomena in horizontal wells. Fuel, 153: 276-283.

Al-Ismael, M., Awotunde, A., Al-Yousef, H. and Al-Hashim, H., 2018. A Well Placement Optimization Constrained to Regional Pressure Balance, SPE Europec featured at 80th EAGE Conference and Exhibition. Society of Petroleum Engineers.

Almeida, L.F., Vellasco, M.M. and Pacheco, M.A., 2010. Optimization system for valve control in intelligent wells under uncertainties. Journal of Petroleum Science and Engineering, 73(1-2): 129-140.

Arouri, Y. and Sayyafzadeh, M., 2020. An accelerated gradient algorithm for well control optimization. Journal of Petroleum Science and Engineering, 190: 106872.

Bangerth, W., Klie, H., Wheeler, M., Stoffa, P. and Sen, M., 2006. On optimization algorithms for the reservoir oil well placement problem. Computational Geosciences, 10(3): 303-319.

Behnke, S., 2003. Hierarchical neural networks for image interpretation, 2766. Springer.

Bouzarkouna, Z., Ding, D.Y. and Auger, A., 2012. Well placement optimization with the covariance matrix adaptation evolution strategy and meta-models. Computational Geosciences, 16(1): 75-92.

Bukshtynov, V., Volkov, O., Durlofsky, L.J. and Aziz, K., 2015. Comprehensive framework for gradient-based optimization in closed-loop reservoir management. Computational Geosciences, 19(4): 877-897.

Cardoso, M.A. and Durlofsky, L.J., 2010. Linearized reduced-order models for subsurface flow simulation. Journal of Computational Physics, 229(3): 681-700.

Chen, C., Wang, Y., Li, G. and Reynolds, A.C., 2010. Closed-loop reservoir management on the Brugge test case. Computational Geosciences, 14(4): 691-703.

Chen, G. et al., 2020. Surrogate-assisted evolutionary algorithm with dimensionality reduction method for water flooding production optimization. Journal of Petroleum Science and Engineering, 185: 106633.

Cheng, K. and Lu, Z., 2020. Structural reliability analysis based on ensemble learning of surrogate models. Structural Safety, 83: 101905.

Chu, M.-g. et al., 2020. Determination of an infill well placement using a data-driven multi-modal convolutional neural network. Journal of Petroleum Science and Engineering, 195: 106805.

Chugh, T., Jin, Y., Miettinen, K., Hakanen, J. and Sindhya, K., 2016. A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization. IEEE Transactions on Evolutionary Computation, 22(1): 129-142.

de Brito, D.U. and Durlofsky, L.J., 2020a. Field development optimization using a sequence of surrogate treatments. Computational Geosciences: 1-31.

de Brito, D.U. and Durlofsky, L.J., 2020b. Well control optimization using a two-step surrogate treatment. Journal of Petroleum Science and Engineering, 187: 106565.

Ding, S., Lu, R., Xi, Y., Liu, G. and Ma, J., 2020. Efficient well placement optimization coupling hybrid objective function with particle swarm optimization algorithm. Applied Soft Computing, 95: 106511.

Ding, S., Lu, R., Xi, Y., Wang, S. and Wu, Y., 2019. Well placement optimization using direct mapping of productivity potential and threshold value of productivity potential management strategy. Computers & Chemical Engineering, 121: 327-337.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A. and Vapnik, V., 1997. Support vector regression machines. Advances in neural information processing systems, 9: 155-161.

Durlofsky, L., 2010. Use of reduced-order modeling procedures for production optimization. SPE Journal, 15(02): 426-435.

Eberhart, R. and Kennedy, J., 1995. A new optimizer using particle swarm theory, MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. Ieee, pp. 39-43.

563   Enab, K. and Ertekin, T., 2021. Screening and optimization of CO2-WAG injection and fish-bone well
564   structures in low permeability reservoirs using artificial neural network. Journal of
565   Petroleum Science and Engineering, 200: 108268.
566   Figueiredo, M.A., 2003. Adaptive sparseness for supervised learning. IEEE transactions on pattern
567   analysis and machine intelligence, 25(9): 1150-1159.
568   Fonseca, R.R.M., Chen, B., Jansen, J.D. and Reynolds, A., 2017. A stochastic simplex approximate
569   gradient (StoSAG) for optimization under uncertainty. International Journal for Numerical
570   Methods in Engineering, 109(13): 1756-1776.
571   Forouzanfar, F. and Reynolds, A., 2014. Joint optimization of number of wells, well locations and
572   controls using a gradient-based algorithm. Chemical Engineering Research and Design, 92(7):
573   1315-1328.
574   Gers, F.A., Eck, D. and Schmidhuber, J., 2002. Applying LSTM to time series predictable through time-
575   window approaches, Neural Nets WIRN Vietri-01. Springer, pp. 193-200.
576   Glorot, X., Bordes, A. and Bengio, Y., 2011. Deep sparse rectifier neural networks, Proceedings of the
577   fourteenth international conference on artificial intelligence and statistics. JMLR Workshop
578   and Conference Proceedings, pp. 315-323.
579   Goel, T., Haftka, R.T., Shyy, W. and Queipo, N.V., 2007. Ensemble of surrogates. Structural and
580   Multidisciplinary Optimization, 33(3): 199-216.
581   Golzari, A., Sefat, M.H. and Jamshidi, S., 2015. Development of an adaptive surrogate model for
582   production optimization. Journal of petroleum Science and Engineering, 133: 677-688.
583   Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y., 2016. Deep learning, 1. MIT press Cambridge.
584   Gouda, A. et al., 2021. Development of an artificial neural network model for predicting the dew
585   point pressure of retrograde gas condensate. Journal of Petroleum Science and Engineering:
586   109284.
587   Guo, Z. and Reynolds, A.C., 2018. Robust Life-Cycle Production Optimization With a Support-Vector-
588   Regression Proxy. SPE Journal.
589   Güyagüler, B., Horne, R.N., Rogers, L. and Rosenzweig, J.J., 2002. Optimization of well placement in a
590   Gulf of Mexico waterflooding project. SPE Reservoir Evaluation & Engineering, 5(03): 229-
591   236.
592   Haghighat Sefat, M., Elsheikh, A.H., Muradov, K.M. and Davies, D.R., 2016. Reservoir uncertainty
593   tolerant, proactive control of intelligent wells. Computational Geosciences, 20(3): 655-676.
594   He, J. and Durlofsky, L.J., 2014. Reduced-order modeling for compositional simulation by use of
595   trajectory piecewise linearization. SPE Journal, 19(05): 858-872.
596   Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.R., 2012. Improving
597   neural networks by preventing co-adaptation of feature detectors. arXiv preprint
598   arXiv:1207.0580.
599   Holland, J.H., 1975. Adaptation in natural and artificial systems, University of Michigan press. Ann
600   arbor, MI, 1(97): 5.
601   Horowitz, B., Afonso, S.M.B. and de Mendonça, C.V.P., 2013. Surrogate based optimal waterflooding
602   management. Journal of Petroleum Science and Engineering, 112: 206-219.
603   Hua, Y. et al., 2019. Deep learning with long short-term memory for time series prediction. IEEE
604   Communications Magazine, 57(6): 114-119.
605   Ioffe, S. and Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing
606   internal covariate shift, International conference on machine learning. PMLR, pp. 448-456.
607   Jain, A.K., Mao, J. and Mohiuddin, K.M., 1996. Artificial neural networks: A tutorial. Computer, 29(3):
608   31-44.
609   Jansen, J.-D. et al., 2014. The egg model–a geological ensemble for reservoir simulation. Geoscience
610   Data Journal, 1(2): 192-195.
611   Jesmani, M., Jafarpour, B., Bellout, M., Hanea, R. and Foss, B., 2016. Application of simultaneous
612   perturbation stochastic approximation to well placement optimization under uncertainty,

613   ECMOR XV-15th European Conference on the Mathematics of Oil Recovery. European
614   Association of Geoscientists & Engineers, pp. cp-494-00133.
615 Jesmani, M., Jafarpour, B., Bellout, M.C. and Foss, B., 2020. A reduced random sampling strategy for
616   fast robust well placement optimization. Journal of Petroleum Science and Engineering, 184:
617   106414.
618 Jiang, S., Sun, W. and Durlofsky, L.J., 2019. A data-space inversion procedure for well control
619   optimization and closed-loop reservoir management. Computational Geosciences: 1-19.
620 Jin, Y., 2005. A comprehensive survey of fitness approximation in evolutionary computation. Soft
621   computing, 9(1): 3-12.
622 Jin, Y., 2011. Surrogate-assisted evolutionary computation: Recent advances and future challenges.
623   Swarm and Evolutionary Computation, 1(2): 61-70.
624 Junior, J.R.B. et al., 2021. A comparison of machine learning surrogate models for net present value
625   prediction from well placement binary data. Journal of Petroleum Science and Engineering:
626   109208.
627 Kahrobaei, S., Van Essen, G., Van Doren, J., Van den Hof, P. and Jansen, J., 2013. Adjoint-based
628   history matching of structural models using production and time-lapse seismic data, SPE
629   Reservoir Simulation Symposium. Society of Petroleum Engineers.
630 Kim, J., Lee, K. and Choe, J., 2021. Efficient and robust optimization for well patterns using a PSO
631   algorithm with a CNN-based proxy model. Journal of Petroleum Science and Engineering:
632   109088.
633 Kim, J., Yang, H. and Choe, J., 2020. Robust optimization of the locations and types of multiple wells
634   using CNN based proxy models. Journal of Petroleum Science and Engineering: 107424.
635 Knowles, J., 2006. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive
636   multiobjective optimization problems. IEEE Transactions on Evolutionary Computation,
637   10(1): 50-66.
638 LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time series. The
639   handbook of brain theory and neural networks, 3361(10): 1995.
640 LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553): 436-444.
641 LeCun, Y. et al., 1989. Backpropagation applied to handwritten zip code recognition. Neural
642   computation, 1(4): 541-551.
643 LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document
644   recognition. Proceedings of the IEEE, 86(11): 2278-2324.
645 Li, L., Jafarpour, B. and Mohammad-Khaninezhad, M.R., 2013. A simultaneous perturbation
646   stochastic approximation algorithm for coupled well placement and control optimization
647   under geologic uncertainty. Computational Geosciences, 17(1): 167-188.
648 Liu, W., Liu, W.D. and Gu, J., 2020. Forecasting oil production using ensemble empirical model
649   decomposition based Long Short-Term Memory neural network. Journal of Petroleum
650   Science and Engineering, 189: 107013.
651 Liu, X. and Reynolds, A.C., 2016. Augmented lagrangian method for maximizing expectation and
652   minimizing risk for optimal well-control problems with nonlinear constraints. SPE Journal,
653   21(05): 1,830-1,842.
654 Liu, Z. and Reynolds, A.C., 2020. A Sequential-Quadratic-Programming-Filter Algorithm with a
655   Modified Stochastic Gradient for Robust Life-Cycle Optimization Problems with Nonlinear
656   State Constraints. SPE Journal.
657 Lu, R., Forouzanfar, F. and Reynolds, A.C., 2017. An efficient adaptive algorithm for robust control
658   optimization using StoSAG. Journal of Petroleum Science and Engineering, 159: 314-330.
659 Lu, R. and Reynolds, A.C., 2020. Joint Optimization of Well Locations, Types, Drilling Order, and
660   Controls Given a Set of Potential Drilling Paths. SPE Journal.
661 Ma, Z. and Leung, J.Y., 2020. Design of warm solvent injection processes for heterogeneous heavy oil
662   reservoirs: A hybrid workflow of multi-objective optimization and proxy models. Journal of
663   Petroleum Science and Engineering: 107186.

664 Maas, A.L., Hannun, A.Y. and Ng, A.Y., 2013. Rectifier nonlinearities improve neural network acoustic
665   models, Proc. icml. Citeseer, pp. 3.
666 McKay, M.D., Beckman, R.J. and Conover, W.J., 2000. A comparison of three methods for selecting
667   values of input variables in the analysis of output from a computer code. Technometrics,
668   42(1): 55-61.
669 Panahli, C., 2017. Implementation of Particle Swarm Optimization Algorithm within FieldOpt
670   Optimization Framework-Application of the algorithm to well placement optimization,
671   NTNU.
672 Panja, P., Velasco, R., Pathak, M. and Deo, M., 2018. Application of artificial intelligence to forecast
673   hydrocarbon production from shales. Petroleum, 4(1): 75-89.
674 Park, H.-Y. and Datta-Gupta, A., 2011. Reservoir management using streamline-based flood
675   efficiency maps and application to rate optimization, SPE Western North American Region
676   Meeting. OnePetro.
677 Peters, L. et al., 2010. Results of the Brugge benchmark study for flooding optimization and history
678   matching. SPE Reservoir Evaluation & Engineering, 13(03): 391-405.
679 Rahmanifard, H. and Plaksina, T., 2019. Application of artificial intelligence techniques in the
680   petroleum industry: a review. Artificial Intelligence Review, 52(4): 2295-2318.
681 Razak, S.M. and Jafarpour, B., 2020. Convolutional neural networks (CNN) for feature-based model
682   calibration under uncertain geologic scenarios. Computational Geosciences, 24(4): 1625-
683   1649.
684 Razavi, S., Tolson, B.A. and Burn, D.H., 2012. Review of surrogate modeling in water resources.
685   Water Resources Research, 48(7).
686 Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1985. Learning internal representations by error
687   propagation, California Univ San Diego La Jolla Inst for Cognitive Science.
688 Sabah, M., Talebkeikhah, M., Agin, F., Talebkeikhah, F. and Hasheminasab, E., 2019. Application of
689   decision tree, artificial neural networks, and adaptive neuro-fuzzy inference system on
690   predicting lost circulation: A case study from Marun oil field. Journal of Petroleum Science
691   and Engineering, 177: 236-249.
692 Sagheer, A. and Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM
693   recurrent networks. Neurocomputing, 323: 203-213.
694 Salehian, M. and Çınar, M., 2019. Reservoir characterization using dynamic capacitance–resistance
695   model with application to shut-in and horizontal wells. Journal of Petroleum Exploration and
696   Production Technology, 9(4): 2811-2830.
697 Salehian, M., Haghighat Sefat, M. and Muradov, K., 2020a. A MultiSolution Optimization Framework
698   for Well Placement and Control SPE Reservoir Evaluation & Engineering.
699 Salehian, M., Haghighat Sefat, M. and Muradov, K., 2020b. A Robust Multi-Solution Framework for
700   Well Location and Control Optimization, ECMOR XVII - 17th European Conference on the
701   Mathematics of Oil Recovery. European Association of Geoscientists and Engineers,
702   Edinburgh, United Kingdom.
703 Sayyafzadeh, M., 2017. Reducing the computation time of well placement optimisation problems
704   using self-adaptive metamodelling. Journal of Petroleum Science and Engineering, 151: 143-
705   158.
706 Scherer, D., Müller, A. and Behnke, S., 2010. Evaluation of pooling operations in convolutional
707   architectures for object recognition, International conference on artificial neural networks.
708   Springer, pp. 92-101.
709 Schlumberger, 2017. ECLIPSE® User Manual.
710 Song, X. et al., 2020. Time-series well performance prediction based on Long Short-Term Memory
711   (LSTM) neural network model. Journal of Petroleum Science and Engineering, 186: 106682.
712 Spall, J.C., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient
713   approximation. IEEE transactions on automatic control, 37(3): 332-341.

714 Spall, J.C., 1998. Implementation of the simultaneous perturbation algorithm for stochastic
715   optimization. IEEE Transactions on aerospace and electronic systems, 34(3): 817-823.
716 Specht, D.F., 1990. Probabilistic neural networks. Neural networks, 3(1): 109-118.
717 Sun, Q. and Ertekin, T., 2020. Screening and optimization of polymer flooding projects using artificial-
718   neural-network (ANN) based proxies. Journal of Petroleum Science and Engineering, 185:
719   106617.
720 Tavallali, M., Karimi, I., Teo, K., Baxendale, D. and Ayatollahi, S., 2013. Optimal producer well
721   placement and production planning in an oil reservoir. Computers & Chemical Engineering,
722   55: 109-125.
723 Tavallali, M.S., Bakhtazma, F., Meymandpour, A. and Karimi, I.A., 2018. Optimal Drilling Planning by
724   Considering the Subsurface Dynamics–Combing the Flexibilities of Modeling and a Reservoir
725   Simulator. Industrial & Engineering Chemistry Research.
726 Thiele, M.R. and Batycky, R., 2003. Water injection optimization using a streamline-based workflow,
727   SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.
728 Trehan, S. and Durlofsky, L.J., 2016. Trajectory piecewise quadratic reduced-order model for
729   subsurface flow, with application to PDE-constrained optimization. Journal of Computational
730   Physics, 326: 446-473.
731 Ushmaev, O.S. et al., 2019. Efficient brownfield optimization of a reservoir in west Siberia. Petroleum
732   Geoscience, 25(2): 207-218.
733 Van Doren, J.F., Markovinović, R. and Jansen, J.-D., 2006. Reduced-order optimal control of water
734   flooding using proper orthogonal decomposition. Computational Geosciences, 10(1): 137-
735   158.
736 Van Essen, G., Van den Hof, P. and Jansen, J.-D., 2011. Hierarchical long-term and short-term
737   production optimization. SPE Journal, 16(01): 191-199.
738 Viana, F.A., Haftka, R.T. and Watson, L.T., 2013. Efficient global optimization algorithm assisted by
739   multiple surrogate techniques. Journal of Global Optimization, 56(2): 669-689.
740 Volkov, O. and Bellout, M.C., 2018. Gradient-based constrained well placement optimization. Journal
741   of Petroleum Science and Engineering, 171: 1052-1066.
742 Wang, H., Echeverría-Ciaurri, D., Durlofsky, L. and Cominelli, A., 2012. Optimal well placement under
743   uncertainty using a retrospective optimization framework. Spe Journal, 17(01): 112-121.
744 Wang, X., Haynes, R.D. and Feng, Q., 2016. A multilevel coordinate search algorithm for well
745   placement, control and joint optimization. Computers & Chemical Engineering, 95: 75-96.
746 Wang, X., Haynes, R.D., He, Y. and Feng, Q., 2019. Well control optimization using derivative-free
747   algorithms and a multiscale approach. Computers & Chemical Engineering, 123: 12-33.
748 Yeten, B., Durlofsky, L.J. and Aziz, K., 2003. Optimization of nonconventional well type, location, and
749   trajectory. SPE Journal, 8(03): 200-210.
750 Zerpa, L.E., Queipo, N.V., Pintos, S. and Salager, J.-L., 2005. An optimization methodology of alkaline–
751   surfactant–polymer flooding processes using field scale numerical simulation and multiple
752   surrogates. Journal of Petroleum Science and Engineering, 47(3-4): 197-208.
753 Zhang, G. et al., 2021. Pattern visualization and understanding of machine learning models for
754   permeability prediction in tight sandstone reservoirs. Journal of Petroleum Science and
755   Engineering, 200: 108142.
756 Zhang, Q., Liu, W., Tsang, E. and Virginas, B., 2009. Expensive multiobjective optimization by
757   MOEA/D with Gaussian process model. IEEE Transactions on Evolutionary Computation,
758   14(3): 456-474.
759 Zhao, D. and Xue, D., 2011. A multi-surrogate approximation method for metamodeling. Engineering
760   with Computers, 27(2): 139-153.
761 Zhao, M. et al., 2020. A surrogate-assisted multi-objective evolutionary algorithm with dimension-
762   reduction for production optimization. Journal of Petroleum Science and Engineering, 192:
763   107192.

764   Zingg, D.W., Nemec, M. and Pulliam, T.H., 2008. A comparative evaluation of genetic and gradient-
765   based algorithms applied to aerodynamic optimization. European Journal of Computational
766   Mechanics/Revue Européenne de Mécanique Numérique, 17(1-2): 103-126.

767