Computers & Geosciences 91 (2016) 19–32

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Case study

# Performance comparison of several response surface surrogate models and ensemble methods for water injection optimization under uncertainty

Masoud Babaei [a,*], Indranil Pan [b]

[a] School of Chemical Engineering and Analytical Science, University of Manchester, M13 9PL, UK
[b] Department of Earth Science & Engineering, Imperial College London, SW7 2AZ, UK

ARTICLE INFO

*Article history:*
Received 8 February 2015
Received in revised form
25 February 2016
Accepted 26 February 2016
Available online 2 March 2016

*Keywords:*
Surrogate modeling
Mixture surrogates
Water injection optimization

ABSTRACT

In this paper we defined a relatively complex reservoir engineering optimization problem of maximizing the net present value of the hydrocarbon production in a water flooding process by controlling the water injection rates in multiple control periods. We assessed the performance of a number of response surface surrogate models and their ensembles which are combined by Dempster–Shafer theory and Weighted Averaged Surrogates as found in contemporary literature works. Most of these ensemble methods are based on the philosophy that multiple weak learners can be leveraged to obtain one strong learner which is better than the individual weak ones. Even though these techniques have been shown to work well for test bench functions, we found them not offering a considerable improvement compared to an individually used cubic radial basis function surrogate model. Our simulations on two and three dimensional cases, with varying number of optimization variables suggest that cubic radial basis functions-based surrogate model is reliable, outperforms Kriging surrogates and multivariate adaptive regression splines, and if it does not outperform, it is rarely outperformed by the ensemble surrogate models.


## 1. Introduction

The inherent uncertain nature of the geological models—due to the sparseness and scarcity of data over vast physical domains—require the designer to assume multiple geological realizations to predict flow. This adds up prohibitive computational costs to already demanding single-realization models. Moreover, the control parameters for the fluid injection and production wells in the industrial scale are often numerous and are subject to operational constraints and time-dependent uncertainties. This also makes the computational domain of the optimization problem large and consequently difficult, if not impossible, to handle for even the modern computing systems. Surrogate models are an attractive option in such circumstances. Surveys of implementation of surrogates for optimization purposes in broad engineering applications can be found in Jin (2005) and Jin (2011). The improvement of computational efficiency of surrogate-based optimization compared to the traditional optimization such as genetic algorithm has been shown in Ong et al. (2003).

The surrogate models are however approximations of the original objective functions, therefore they might introduce artificial optimal solutions which do not exist in the original objective function (Jin, 2011) and lead to premature convergence. Also the techniques have shown a strong dependence on the complex dynamics of the non-linear interactions in the model, dimension of the design space, etc. (Zubarev, 2009). Therefore a proper surrogate model management strategy is very important (Jin, 2011). The choice of a particular surrogate model is also problem-dependent and for a given problem, it is not trivial to decide which surrogate model would give the best optimization result. It has been shown that one surrogate model might give good results for a particular problem while it might perform very poorly when applied to another problem (Viana and Haftka, 2008). One solution to these shortcomings might be to improve the accuracy of the surrogates so that they are less prone to over-fitting and have more generalization capabilities for unseen solution points (Jin et al., 2002).

Another approach can be the use of multiple surrogate models (Goel et al., 2007) which have been shown to be beneficial from the optimization point of view. In this line of research, Zhou et al. (2007) employed multiple surrogates such as regression and interpolating local surrogates to provide a diversity of approximation models in a multi-surrogates assisted memetic algorithm.

* Corresponding author.
*E-mail addresses:* masoud.babaei@manchester.ac.uk (M. Babaei), i.pan11@imperial.ac.uk (I. Pan).

http://dx.doi.org/10.1016/j.cageo.2016.02.022
0098-3004/

Gorissen et al. (2009) brought multiple surrogates to adaptive sampling. The objective is to be able to select the best surrogate model by adding points iteratively. Glaz et al. (2009) implemented a weighted-average approach in use of multiple surrogates. In their application they have found that at relatively little additional cost compared with optimizing with a single surrogate, multiple surrogates can be used to locate extrema of the objective function of their interest that would be overlooked if only a single approximation method was employed. Zhang et al. (2012) developed a hybrid surrogate modeling methodology that adaptively combines the favorable characteristics of different surrogate models including RBF and Kriging. The methodology generates different surrogate models (component surrogates), and weights aggregation of the estimated function value based on the local measure of accuracy of the individual surrogates.

In an attempt to put into test some of the developments that have shown to work mostly for analytically tractable problems and test bench functions, we assess the performance of two open source toolboxes which use ensemble surrogate strategies. These toolboxes are due to Müller and Piché (2011) that uses Dempster–Shafer theory to mix surrogate models, and Viana et al. (2013) that uses the multiple surrogates based on the square root of the prediction sum of squares for surrogate selection. These methods have not been used in the context of geo-engineering. So we assess the performance of these developments of uncertainty-laden models of heterogeneous reservoirs that adds to the complexity of the models. Also we apply two different optimization strategies to search for the optimal solution.

## 2. Model description, uncertainty and optimization problem

### 2.1. Governing equations

The water injection process into the oil reservoir is considered herein with assumptions of an immiscible and incompressible multiphase fluid flow with unit formation volume factor for oil and water. Gravity and capillary effects are neglected. The problem is described and sequentially solved by Darcy's law ($v_t = -\mathbf{K}\lambda_t(S_w)\nabla p$ in $\Omega$), mass conservation equation ($\nabla\cdot v_t = q$) and the transport equation ($\varphi\frac{\partial S_w}{\partial t} + \nabla\cdot[v_t f_w(S_w)] = q_w$), where $v_t = v_o + v_w$ is the total Darcy velocity [m$^3$/day] of the engaging fluids (oil and water phases denoted by subscripts $o$ and $w$, respectively), $q$ represents the volumetric total source and sink contributions [m$^3$/day] of oil and water phases from the wells and boundary conditions. Also, $\mathbf{K}$ [mD] is the tensor of absolute permeability, $\lambda_t(S_w) = \lambda_o(S_w) + \lambda_w(S_w)$ is the total mobility and is a function of water saturation, $S_w$. The fluid pressure, $p$ [atm], is, in the absence of capillarity, equal to oil and water phase pressures, $p_o$ and $p_w$. Finally $\Omega$ is the problem domain.

In the transport equation, $\varphi$ [−] is the porosity of the porous medium, $f_w(S_w)$ [−] is the fractional flow function of water defined by $f_w(S_w) = \lambda_w/\lambda_t$. The phase mobilities ($\lambda_o$ and $\lambda_w$) are herein modeled by polynomial water and oil relative permeability curves, $k_{ro}(S_{wD}) = k_{ro,\max}(1 - S_{wD})^{n_o}$ and $k_{rw}(S_{wD}) = k_{rw,\max}(S_{wD})^{n_w}$ and constant phase viscosities, $\mu_o$ and $\mu_w$, as $\lambda_w(S_w) = k_{rw}(S_{wD})/\mu_w$ and $\lambda_o(S_w) = k_{ro}(S_{wD})/\mu_o$, where $S_{wD} = (S_w - S_{wc})/(1 - S_{or} - S_{wc})$. In these relations $n_o$ and $n_w$ are exponents of the polynomials controlling curvature of the curves, $S_{wD}$ is the normalized water saturation that varies between zero and one as opposed to the water saturation that varies between $S_{wc}$ (connate water saturation) and $1 - S_{or}$ where $S_{or}$ is the oil residual saturation.

The above equations are solved with the open-source MATLAB Reservoir Simulation Toolbox (Lie et al., 2012).

### 2.2. Geological model

#### 2.2.1. Two dimensional model

The two dimensional geological model used in this work is 3000 m × 3000 m × 1 m long in $x$, $y$ and $z$ directions representing a thin horizontal reservoir. The gridblocks are 50 m × 50 m × 1 in length, width and height respectively so that the number of gridblocks is 60 × 60. The boundaries are assumed fully closed and the reservoir is fully saturated with oil. The porosity of the model is a constant value of 0.2. The water and oil viscosities are $1.0 \times 10^{-3}$ Pa s and $10.0 \times 10^{-3}$ Pa s. The water and oil surface densities are 1014 and 859 kg m$^{-3}$. The relative permeabilities of oil and water are represented by quadratic polynomials ($n_o = n_w = 2$) and $k_{ro,\max} = k_{rw,\max} = 1$ and capillary pressure is ignored and the initial water saturation is set to zero.

The permeability is assumed uncertain but exhibiting, in two separate cases, the features of either of the following geo-environmental landscapes: a shale-dominant reservoir with multiple narrow diagonal intersecting channels with 45° orientation (denoted hereafter simply by Model 2D-a) and a sandstone reservoir crisscrossed with a multitude of lateral shale streaks (Model 2D-b).

In order to generate realizations of different permeability fields, S-GeMS (the Stanford Geostatistical Modeling Software available at http://sgems.sourceforge.net) is used. S-GeMS provides algorithms for multiple-point geostatistics. A review of multiple-point geostatistics is conducted by Hu and Chugunova (2008) and there are numerous subsurface modeling applications of it in literature (e.g., Ronayne et al., 2008; Mariethoz et al., 2010; Mariethoz and Kelly, 2011). One such algorithm is FILTERSIM (Zhang et al., 2006; Wu et al., 2008) that is used to build the image or numerical model by conditioning to local data patterns using a prior structural model given under the form of a visually explicit training image (Zhang et al., 2006). Reproducing geological shapes based on a training image by multiple-point geostatistics is more realistic than the traditional two-point geostatistics that utilizes variogram models to characterize the spatial structure of data as the variograms often cannot capture curvilinear structures and shapes of geological bodies such as channels (Journel, 1993; Strebelle, 2000).

The training image here is an image of a diagonally channelized permeability field (Model 2D-a) or a shale populated sandstone (Model 2D-b). The training image serves as prior knowledge of the geology of the reservoir. Figs. 1 and 2 show the six realizations of the absolute permeability obtained by unconditional continuous FILTERSIM simulation using S-GeMS for the two cases of permeability considered in this work to introduce uncertainty.

The water injection is performed by four injection wells (I1, I2, I3 and I4) at the corners of the reservoir and one production well in the center of the reservoir (P1) as shown in Fig. 1.

#### 2.2.2. Three dimensional model

We use an ensemble version (with 100 realizations) of the Egg Model (Jansen et al., 2013) for the three dimensional example and we refer to it as Model 3D. The model has 60 × 60 × 7 grid cells of which 18,553 cells are active leaving an egg-shaped model after eliminating the inactive cells. The gridblocks are 8 m × 8 m × 4 m in length, width and height respectively. The porosity is 0.2. Oil and water viscosities are $5.0 \times 10^{-3}$ Pa s and $1.0 \times 10^{-3}$ Pa s. The water and oil surface densities are 1000 and 900 kg m$^{-3}$. For the relative permeabilities of oil and water, $n_o = 4$, $n_w = 3$ and $k_{ro,\max} = 0.8$, $k_{rw,\max} = 0.75$ and $S_{wc} = 0.2$, $S_{or} = 0.1$. Capillary pressure is ignored and the initial water saturation is set 0.1.

The permeability fields (Fig. 3a) demonstrate channelization with strong vertical correlation. There are no aquifer or gas cap in the model, the primary production is neglected, and the production

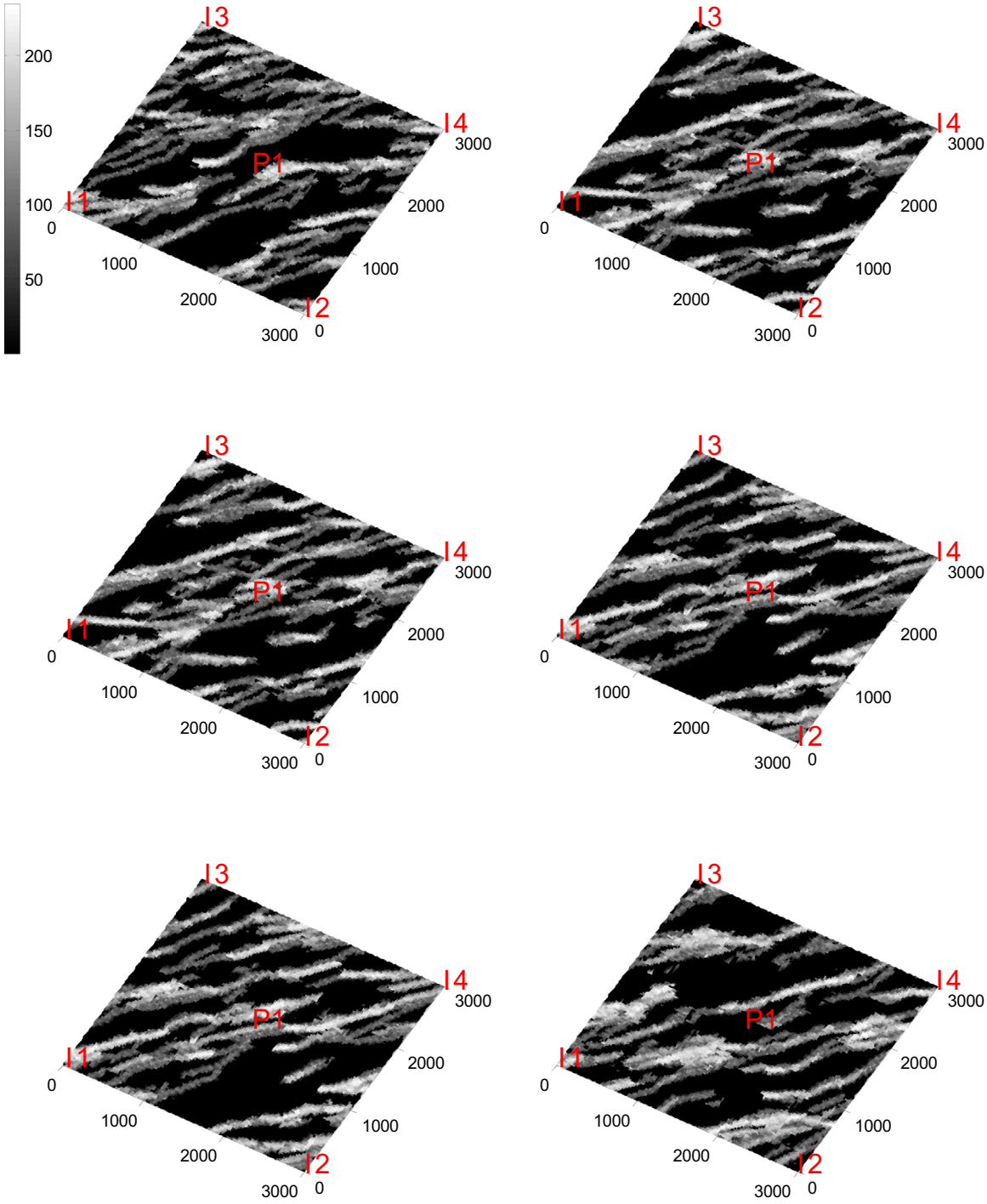


**Fig. 1.** The six realizations of the absolute permeability [mD: millidarcy] for permeability Model 2D-a considered in this study. The realizations exhibit minimum and maximum values of respectively $10^{-3}$ mD (shale, shown in black) and 235 mD (sand, shown in white) for all the realizations.

mechanism is water flooding with eight injection wells (shown in blue in Fig. 3b) and four production wells (shown in red in Fig. 3b).

### 2.3. Optimization problem formulation

The optimization problem is to find a set of optimal water injection rates for the injection wells to maximize the Net Present Value (*NPV*). The objective function is a simple *NPV* model based on the discounted accumulated net cash-flow. A simplified version for two-phase flow (not taking into account installation costs, shut down, etc.) can be expressed as:

$$NPV = \sum_{n=1}^{N}\left[\sum_{j\in \boldsymbol{P}}\left(r_o q_{o,j}^n - r_w q_{w,j}^n\right) - \sum_{j\in \boldsymbol{I}}\left(r_{inj} q_{inj,j}^n\right)\right]\frac{\Delta t^n}{(1+d)^{t_n}} \quad (1)$$

where $N$ is the number of time-steps, $\Delta t$ is the time interval, $\boldsymbol{P}$ and

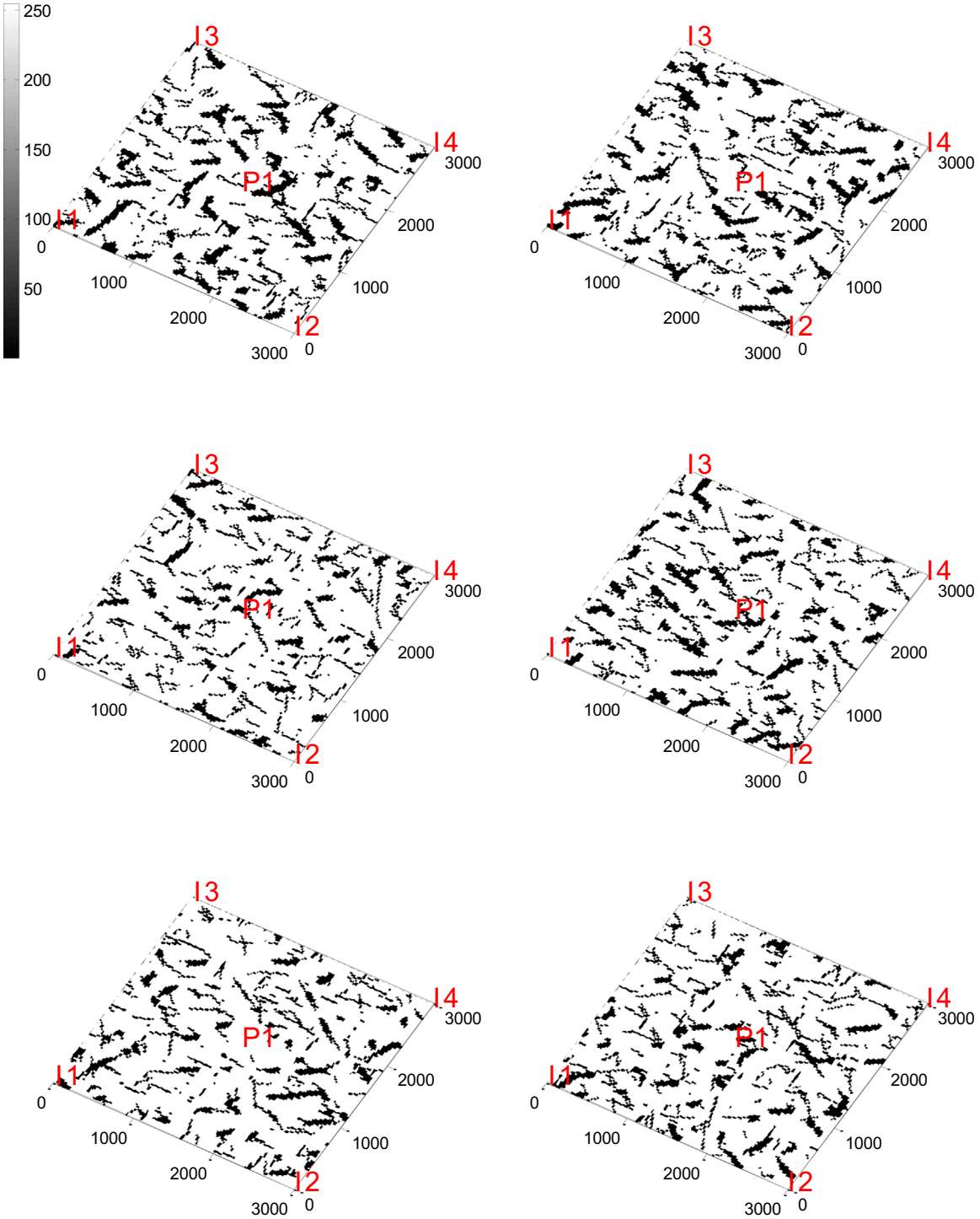


**Fig. 2.** The six realizations of the absolute permeability [mD] for permeability Model 2D-b considered in this study. The realizations exhibit minimum and maximum values of $10^{-3}$ mD (shale, shown in black) and 235 mD (sand, shown in white) respectively for all the realizations.

$\boldsymbol{I}$ are respectively the set of producers and injectors, $q^n_{o,j}$ and $q^n_{w,j}$ [$m^3$/day] the (surface volume) field production rates respectively of oil and water for production well $j$ at time step $n$, $q^n_{inj,j}$ [$m^3$/day] is the field injection rate of water for injection well $j$ at time step $n$, and $r_o$, $r_w$ and $r_{inj}$ [USD/$m^3$] are the oil production revenue, water production cost and water injection cost per volume injected or produced, respectively. Finally, $d$ is the discount rate. The objective function for an ensemble of realizations, to be maximized, is formulated as:

$$
\begin{aligned}
F(\mathbf{u}) &= \frac{1}{N_r}\sum_{\chi=1}^{N_r} F^{\chi}_{NPV}(\mathbf{u}),\ F^{\chi}_{NPV}(\mathbf{u}) \\
&= \sum_{n=1}^{N}\left[\sum_{j\in \boldsymbol{P}}\left(r_o q^{\chi,n}_{o,j} - r_w q^{\chi,n}_{w,j}\right)\right]\frac{\Delta t^n}{(1+d)^{t_n}} \\
&\quad - \sum_{i=1}^{N_p}\left[\sum_{n=n_0^{P_i}}^{n^{P_i}}\left[\sum_{j\in \boldsymbol{I}}\left(r_{inj}u_j^{P_i}q^{\max}_{inj}\right)\right]\frac{\Delta t^n}{(1+d)^{t_n}}\right],
\end{aligned}
\tag{2}
$$

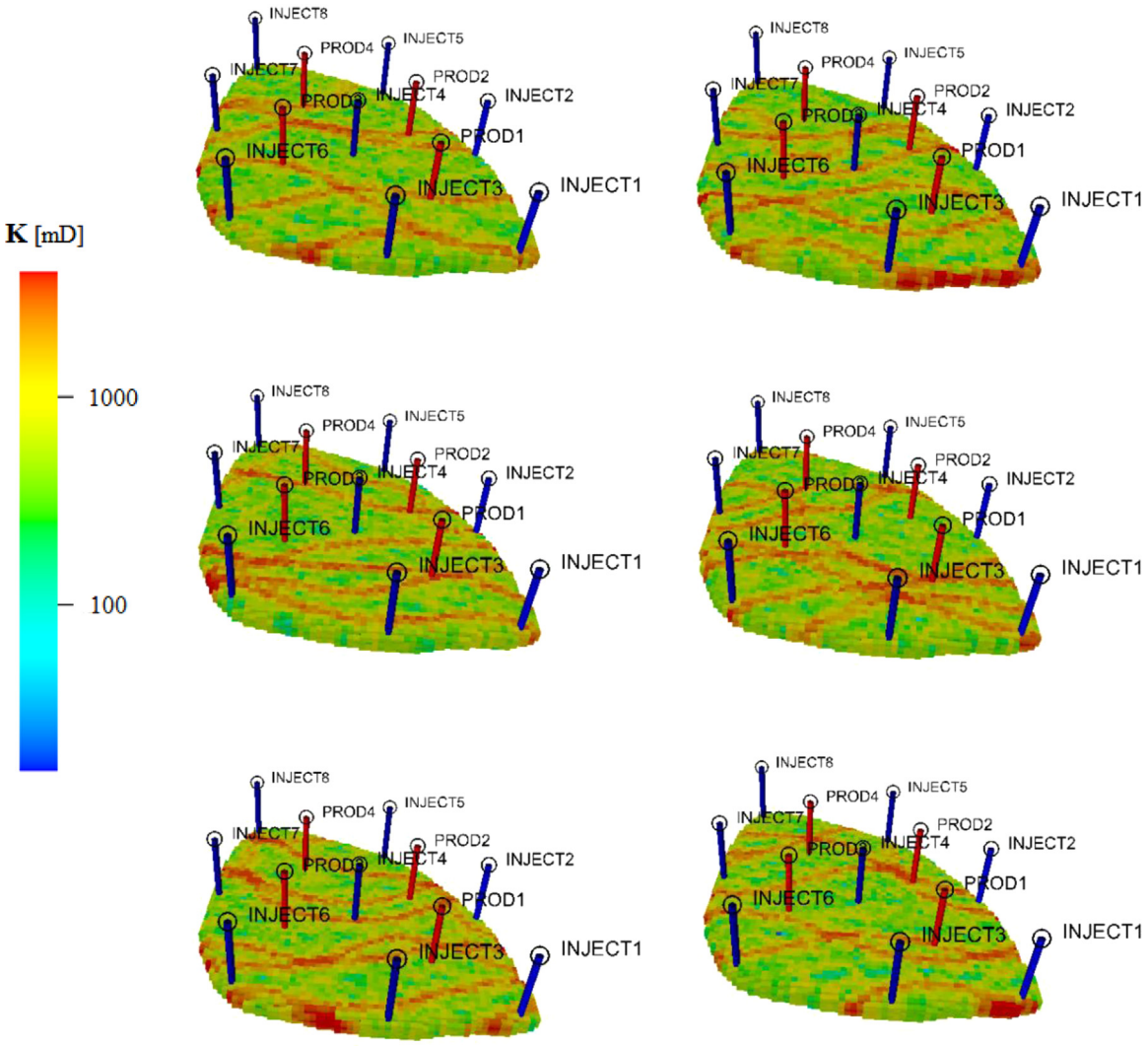


**Fig. 3.** Six chosen realizations of the Egg Model showing the profile of the absolute permeability and the location of injection and production wells. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

where the superscript $\chi$ denotes realization number and $N_r$ is the total number of realizations. $N_p$ is the number of control periods at the start of which the injection rates are modified. Therefore the second term of $F^{\chi}_{NPV}(\mathbf{u})$ is the summation over all periods (denoted by $P_i$, starting from time step $n_0^{P_i}$ and ending at time step $n^{P_i}$) of the cost of injection throughout the simulation time. The maximum allowable or available water injection rate throughout the simulation is denoted by $q^{\max}$. The state variables $\mathbf{u} = \{u_j\}, j \in \boldsymbol{I}$, are the injection multipliers to the maximum allowable injection rate and are updated for each period. The constraints of optimization are set as $0 \le u_j \le 1$, which means that each injection well can inject water at the maximum allowable or available injection rate when the corresponding state variable (injection multiplier) is equal to one or a well can be shut when its corresponding state variable (injection multiplier) is equal to zero.

The number of variables will be the number of injection wells multiplied by the number of control periods. Depending on the number of optimization control parameters, the higher the number of control periods, the more effective the optimization i.e. a higher value of *NPV* would be obtained. This is due to more degrees of freedom in the optimization problem and the algorithm would be able to find a combination of optimization variables which can give a better value of the *NPV*. More number of optimization variables however implies an increase in the number of iterations and consequent higher computational costs for the optimization algorithm to converge.

It should also be noted that the definition as above accounts for the uncertainty of the model by taking the expected value of ensemble of *NPV*'s and finding a solution that works well on average for all the realizations of the model. This is in accordance with the recommendation of van Essen et al. (2009) who proposed to optimize the expectation of *NPV*'s over a set of reservoir models in order to reduce the risk arising from uncertainty in the geological description.

## 3. Combining surrogate models

In contemporary literatures there are numerous implementations of combined/mixed/multiple surrogate models with the promise of achieving better predictions of the true underlying function and obtaining more reliable optimization solutions compared to use of individual surrogate models Bishop (1995), Zhou et al. (2007), Goel et al. (2007), Viana et al. (2009), and Müller and Piché (2011). Out of these publications, we have chosen the last two, which are the most recent ones and have their codes in an open source repository, to be tested on a black-box optimization such as the one we defined in the previous section. Both

algorithms attempt to find a suitable set of weights for individual surrogate models in a mixture surrogate model, which is updated dynamically over the iterations of the optimization algorithm:

$$\hat{y}_{mix}(\mathbf{x}) = \sum_{i=1}^{N_m} w_i \hat{y}_i(\mathbf{x}), \quad \sum_{i=1}^{N_m} w_i = 1, \quad w_i \geq 0 \tag{3}$$

where $\hat{y}_{mix}$ represents the predicted values of the $i^{th}$ surrogate model, $\hat{y}_i$, for the solution vector $\mathbf{x}$, $w_i$ is the weight assigned to the corresponding model and $N_m$ is the number of surrogate models. In the following we briefly describe the two algorithms.

### 3.1. Weighted average surrogates (WAS) approach

Viana et al. (2009) devised a weighted average surrogate of $N_m$ models. The weights were calculated either heuristically, or by the *optimal weighted surrogate* algorithm from minimization of the mean square error of data points at which the surrogate is fitted to the actual model. In both approaches, first the prediction sum of squares ($\mathbf{e}$) is calculated. Then for $p$ data points $\mathbf{e}_{RMS} = \sqrt{\frac{1}{p}\mathbf{e}^T\mathbf{e}}$ is calculated. One way of calculating a vector of $\mathbf{e}$ is calculating the cross-validation error from all data points except the data point where the corresponding entry of $\bar{\mathbf{e}}$ is being calculated. Next, in the heuristic computation of the weights, the algorithm proposed by Goel et al. (2007) is used to calculate

$$w_i = \frac{w_i^*}{\sum_{i=1}^{N_m} w_i^*}, \qquad w_i^* = \left( e_i + \frac{\alpha}{n} \sum_{i=1}^{N_m} e_i \right)^{\beta} \tag{4}$$

where $e_i$ is given by the $\mathbf{e}_{RMS}$ of the $i^{\text{th}}$ surrogate. The two parameters $\alpha$ and $\beta$, control the importance of averaging and of individual $\mathbf{e}$, respectively. Goel et al. (2007) suggested $\alpha = 0.05$ and $\beta = -1$.

For the second approach using proposal of Bishop (1995), the mean square error of a weighted average surrogate is minimized:

$$\arg\min_{\mathbf{w}} MSE = \mathbf{w}^T\mathbf{C}\mathbf{w} \quad \text{subject to} \quad \mathbf{1}^T\mathbf{w} = 1 \tag{5}$$

where $\mathbf{C}$ is the covariance matrix in Bishop's formulation:

$$\mathbf{w}^T\mathbf{C}\mathbf{w} = \frac{1}{V}\int_V \left( y(\mathbf{x}) - \hat{y}_{mix}(\mathbf{x}) \right)^2 d\mathbf{x}$$

$$c_{ij} = \frac{1}{V}\int_V \left( y_i(\mathbf{x}) - \hat{y}_{mix}(\mathbf{x}) \right)\left( y_j(\mathbf{x}) - \hat{y}_{mix}(\mathbf{x}) \right) d\mathbf{x} \tag{6}$$

Alternatively, $c_{ij}$ which is the element $i$ and $j$ for surrogate models $i$ and $j$ can be approximated as $c_{ij} \sim \frac{1}{p}\mathbf{e}_i^T\mathbf{e}_j$ Viana et al. (2009). Using various analytically tractable optimization problems, Viana et al. (2009) concluded that using $\mathbf{e}_{RMS}$ for identifying the best surrogate is a good strategy with increasing number of points.

### 3.2. Dempster–Shafer theory (DST)-based mixed surrogate approach

Müller and Piché (2011) used Dempster–Shafer's theory of evidence as a means of combining information from different sources and to construct a degree of belief. In brief the theory tries to assign the so-called basic probability assignments (BPA's) to different sets that are calculated from evidences – in this case, individual surrogate model's scaled good/bad characteristics such as

1. High/low positive correlation coefficients (CC) between the observed and predicted function values.
2. Low/high root mean squared errors (RMSE) between surrogate response and true function values: $\sqrt{\sum_p (y(x_p) - \hat{y}(x_p))}$,
3. Low/high maximal absolute errors (MAE): $\max(|y(\mathbf{x}) - \hat{y}(\mathbf{x})|)$
4. Low/high median absolute deviation (MAD): $\text{median}(|y(\mathbf{x}) - \hat{y}(\mathbf{x})|)$.

For estimating the belief functions of each surrogate model these four model characteristics are scaled so that the sum of each BPA over all the models is equal to one. The additional constraint that the BPA's cannot be negative Müller and Piché (2011) is also imposed. DST is then applied based on these belief functions to decide the weights assigned to each of the models.

Based on 13 deterministic test cases, an application problem that deals with groundwater bioremediation (Yoon and Shoemaker, 1999) and an application that arises in energy generation using tethered kites, Müller and Shoemaker (2014) concluded that surrogate model mixtures containing radial basis functions model work better than other mixture options.

## 4. Sampling strategy and optimization algorithm

For continuous black-box optimization problems such as the one defined in Section 2, we have used a gradient-free, population-based algorithm that decides on the next expensive function evaluation sampling point by considering both criteria of where the optimal solutions are likely to be located and where the fidelity of the surrogate model can be improved by sampling in unexplored regions of the parameter space. To this end, we use the *candidate sampling* strategy based on Müller and Shoemaker (2014). This is a randomized approach for finding sampling points which is shown to be more successful than optimizing an auxiliary function (Regis and Shoemaker, 2007). The algorithm starts with generating two groups of points, one by uniformly selecting points from the variable domain $D$, and one by perturbing the already sampled points with the lowest function value ($\mathbf{x}_{best} = \arg\min_{\mathbf{x}\in D} f(\mathbf{x})$) with a designated probability, so that $\mathbf{x}_{best}$ are perturbed by randomly adding or subtracting small, medium, or large random perturbations.

Next, the distance of every candidate point to the set of already sampled points and also the objective function value predicted by the surrogate model are considered as two criteria with which the candidate points are scored. The candidate point with the best score is chosen as the next sampling site. A more weight on the distance criterion allows a more global search for unexplored areas of the domain, whereas a more weight on the criterion of the objective function value allows a more local search in the vicinity of promising points as shown in Müller and Shoemaker (2014). To have a balance of both local and global searches, the algorithm implements a cyclic weighting pattern for the two criteria of sampling. It means that the algorithm starts with a large weight to the distance criterion and gradually decreases it to give a proportionally larger weight to the second criterion, and after the weight of the distance criterion has reached its minimum, it is reinitialized and decreased again. With the aid of this algorithm, no auxiliary sub-problem has to be optimized to find the next sample site, and the potential problem of getting trapped in a local optimum of the response surface is avoided as discussed in Müller and Shoemaker (2014). The flowchart of optimization assisted by Candidate Point Strategy and surrogate models is shown in Fig. 4.

In order to be able to rank the different surrogate models and their combinations, and to avoid confusing the effects of a probably outperforming or underperforming optimization algorithm with the effects of using different surrogate models, we use the strategy described above unchanged for all the models. For example, Viana et al. (2013) have used multiple surrogate efficient global optimization algorithm which adds several (as many as the

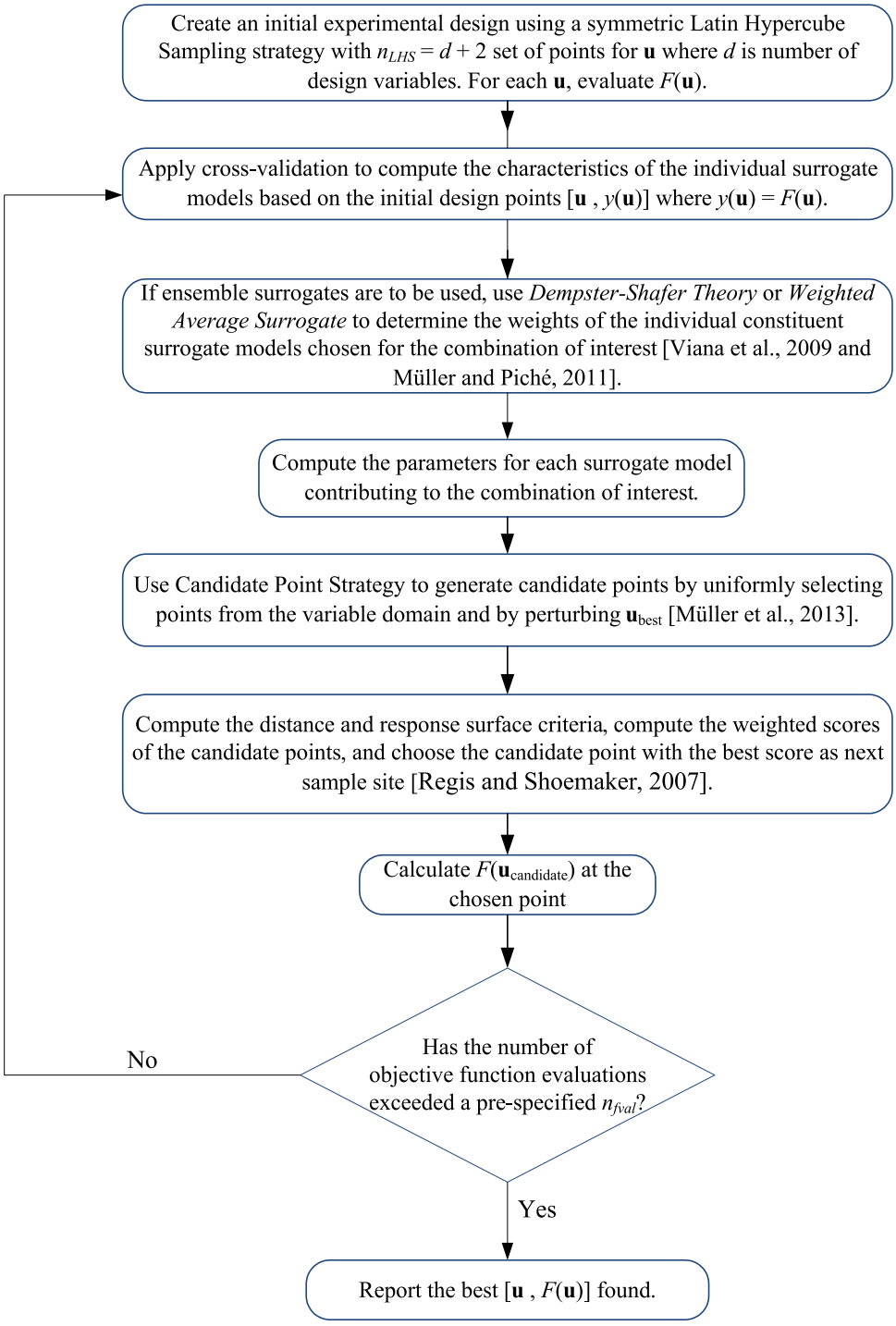


**Fig. 4.** The flowchart of the Candidate Point Strategy for optimization.

surrogate models) points per optimization cycle. This algorithm is suitable when the main concern is wall-clock time rather than number of function evaluation. However, for the sake of comparison we restrict the number of function evaluations, and therefore we use the candidate sampling strategy with WAS as well.

## 5. Optimization results

In this work we consider some of the popular individual and combined surrogate models as also used in the comparisons of Viana et al. (2009) and Müller and Piché (2011). The individual and combined surrogates are listed in Table 1. The selection is based on preliminary results that are not shown in this paper. We selected the high performing individual/combined surrogate models among the available options. For example, since we had observed a very poor performance of polynomial response surface, we have not used it or its mixed models in this comparative work.

### 5.1. Results of optimization on the two-dimensional model

The maximum allowable water injection rate, $q_{inj}^{max}$, is set to 0.001 of pore volume of the reservoir per day which is $4.5 \times 10^3$ m$^3$/day. The injection continues for 2000 days. In three

**Table 1**
List of the individual and combined/multiple surrogate models used in this study.

| Surrogate model | Model number (used in analysis) | Method of combining | Abbreviation |
|---|---|---|---|
| Cubic radial basis function | 1 | – | R |
| Gaussian Kriging | 2 | – | K |
| Multivariate adaptive regression splines | 3 | – | M |
| Cubic radial basis function + Multivariate adaptive regression splines | 4 | DST | RM DST |
| Gaussian Kriging + Multivariate adaptive regression splines | 5 | DST | KM DST |
| Cubic radial basis function + Gaussian Kriging | 6 | DST | RK DST |
| Cubic radial basis function + Gaussian Kriging + Multivariate adaptive regression spline | 7 | DST | RKM DST |
| Gaussian Kriging + Cubic radial basis function | 8 | WAS | RK WAS |
| Gaussian Kriging + Cubic radial basis function + polynomial response surface + linear Shepard model + support vector regression model | 9 | WAS | RKPSV WAS |

separate control-period cases, we assume the number of control periods ($N_p$) to be 1, 2 and 3 to provide 4, 8 and 12 optimization variables respectively. In the first case the rates are not updated, in the second case, the rates are updated on day 1000, in the third case the rates are updated twice on days 667 and 1333. The economic constants are set as $r_o$ = 100 USD/STB, $r_w$ = 10 USD/STB and $r_{inj}$ = 10 USD/STB, where STB stands for stock tank barrel and is equivalent to 0.159 $m^3$. We assume no discount factor ($d = 0$).

The number of function evaluations for Latin hypercube sampling ($N_{LHS}$) and the total number of function evaluations ($N_{eval}$) are specified by the user. We set $N_{LHS}$ equal to 10 multiplied by the number of control periods (10, 20 and 30 for the three control-period cases) and $N_{eval}$ equal to 25 multiplied by the number of control periods (25, 50 and 75 for the three control-period cases).

We ran the optimization problem with each individual or mixed surrogate model 10 times so that the comparisons between the models are fair. In order to be able to run the models 10 times each, we have parallelized the original codes of Müller and Piché (2011), both in calculating the objective function between the 6 realizations and in running each of the 10 runs independently.

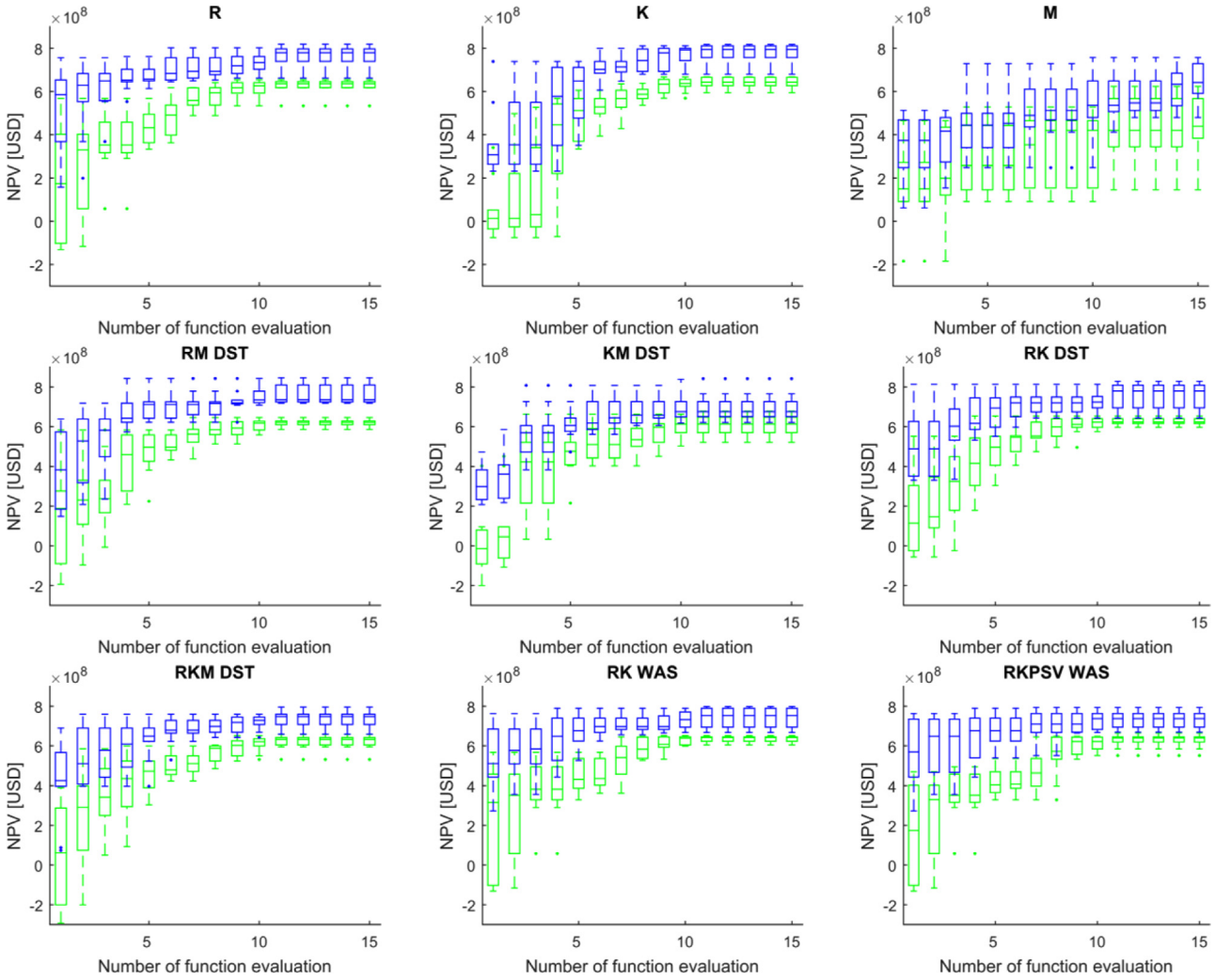


**Fig. 5.** The best solutions obtained per optimization cycles (1 to $N_{eval}$–$N_{LHS}$) by various individual and mixed surrogate models for Model 2D-a (green) and Model 2D-b (blue) using $N_p$=1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

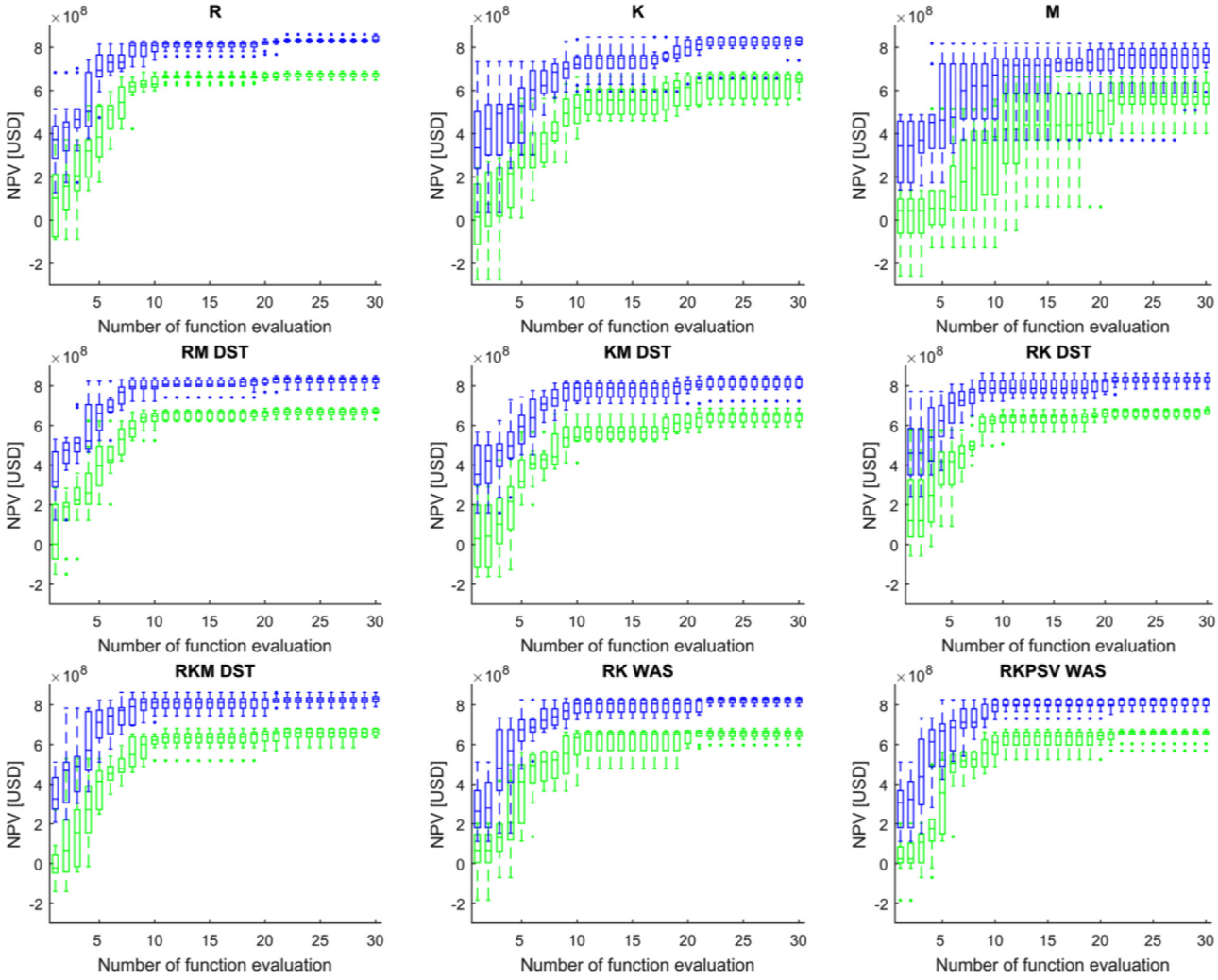


**Fig. 6.** The best solutions obtained per optimization cycles (1 to $N_{eval}$–$N_{LHS}$) by various individual and mixed surrogate models for Model 2D-a (green) and Model 2D-b (blue) using $N_p$=2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The results of the optimization problems for Model 2D-a (channelized) and Model 2D-b (shale) cases, for $N_p$=1, 2 and 3, are shown in the format of box plots at each step of the optimization in Figs. 5–7. We show the results for $F(\mathbf{u})$ after the initial design by Latin hypercube sampling only.

In order to quantitatively interpret the results shown in the above figures, we define the following metrics:

*Metric 1:* The median of the optimization objective function values using a particular surrogate model, $\ell$, between 10 runs of the problem and after the final optimization cycle (25 times the number of control periods). We denote this metric which corresponds to the circles at the last function evaluation point for each surrogate model by $\bar{f}_{NPV,\ell}$ [USD].

*Metric 2:* The best of the optimization objective function values for 10 runs of the problem, denoted by $f^{best}_{NPV,\ell}$ [USD], after the final optimization cycle (25 times the number of control periods).

*Metric 3:* The convergence of the optimization with a particular surrogate model, $\ell$, to $f^{best}_{NPV,\ell}$ at each optimization cycle, *i*, defined by sum of the difference between the best and the worst *NPV* values for iterations after *LHS* sampling stage. The lower this value, the faster the optimization based on the particular surrogate has converged to the best solution. It is an important metric because it shows which models would have performed reliably well if the maximum number of function evaluations was even lower.

The analysis of the results with these metrics is shown in Fig. 8. From the figure we can interpret:

1. Between the individual models, radial basis function surrogate model ($\ell$=1 on the *x*-axis) outperforms Kriging ($\ell$=2) and multivariate adaptive regression splines ($\ell$=3) for almost all the cases and metrics except for metrics 1 and 2 when $N_p$=1 where Kriging is slightly better than radial basis function. However, by increasing the number of control variable to $N_p$=2 or $N_p$=3, radial basis function has outperformed Kriging. This shows that it is more reliable to use radial basis function than Kriging for high number optimization variables. Likewise, in terms of Metric 3 (convergence), radial basis function is considerably superior to Kriging and multivariate adaptive regression splines for $N_p$=2 and 3 as well.
2. None of the mixed surrogate models ($\ell$=4, 5, 6, 7, 8 and 9) exhibits any better, performance (they are sometimes worse off) than radial basis function surrogates.
3. The mixed surrogates are only noticeably useful when their performances are compared with the performance of individual Kriging or multivariate adaptive regression splines. This is particularly manifested by comparing the convergence (Metric 3) for $N_p$=2 and $N_p$=3 of Kriging ($\ell$=2) or multivariate adaptive regression splines ($\ell$=3) respectively with mixed RK ($\ell$=6) and RM ($\ell$=4).

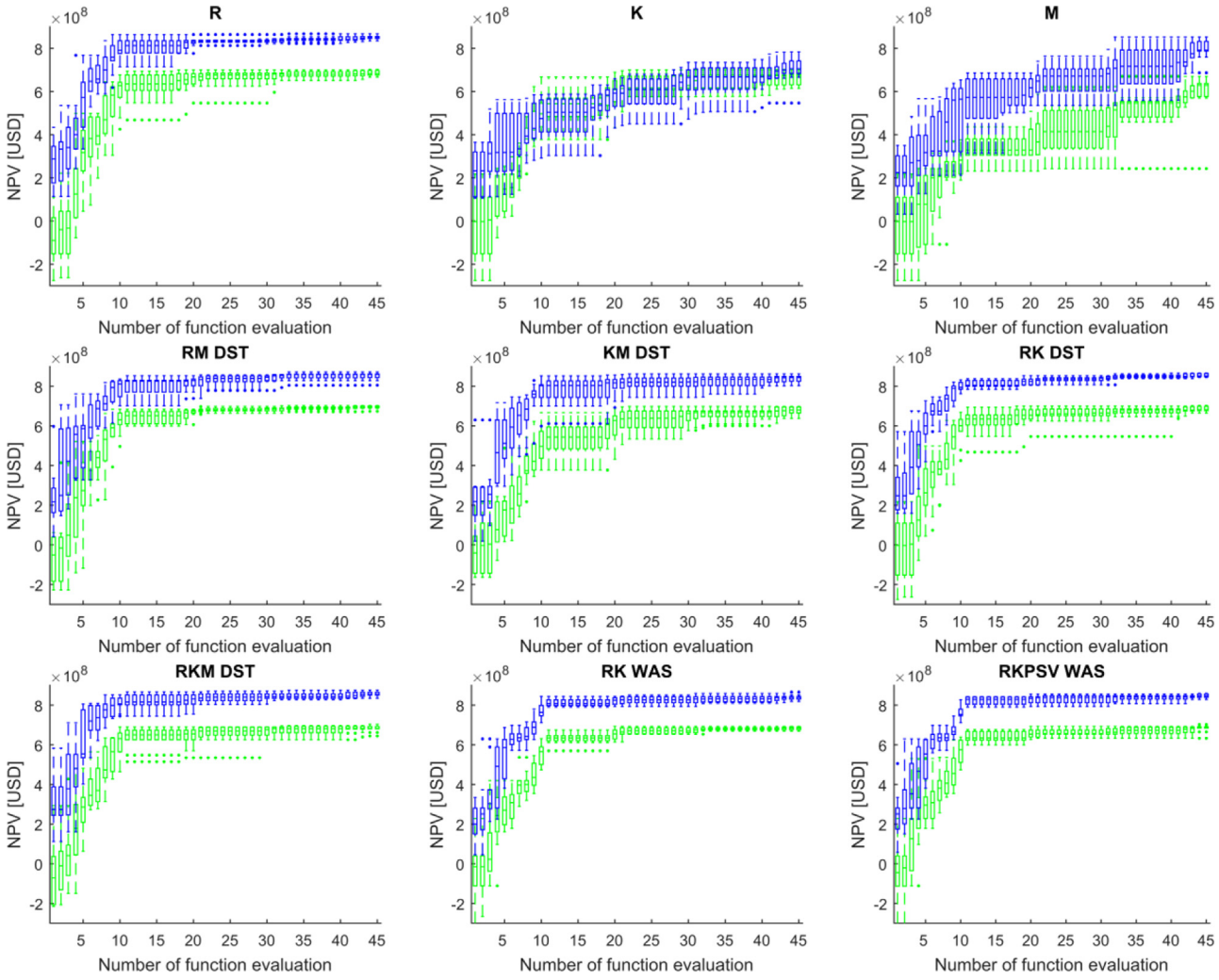


**Fig. 7.** The best solutions obtained per optimization cycles (1 to $N_{eval}$–$N_{LHS}$) by various individual and mixed surrogate models for Model 2D-a (green) and Model 2D-b (blue) using $N_p$=3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The best results for Model 2D-a (channelized system) is *NPV* around $7.03 \times 10^8$ USD, using $N_p$=3. One solution obtained by RK DST with the injection multipliers $\mathbf{u}=10^{-3} \times$[0, 87, 1, 0, 33, 287, 379, 0, 0, 29, 81, and 160]. Given that $q_{inj,j}^{n,\max}$ is $4.5 \times 10^3$ m$^3$/day, the rates of injection for four wells and in three injection periods are $q_{inj}$=[0, 391, 4, 0, 148, 1291, 1705, 0, 0, 130, 364, 72] m$^3$/day.

The best results for Model 2D-b (shale system) is *NPV* around $8.72 \times 10^8$ USD, using $N_p$=3. One solution is obtained by RM DST with the injection multipliers $\mathbf{u}=10^{-3} \times$[ 0, 0, 0, 128, 0, 0, 21, 0, 256, 419, 279, 54]. Given that $q_{inj,j}^{n,\max}$ is $4.5 \times 10^3$ m$^3$/day, the rates of injection for four wells and in three injection periods are $q_{inj}$=[0, 0, 0, 578, 0, 4, 93, 0, 1154, 1886, 1257, 243] m$^3$/day.

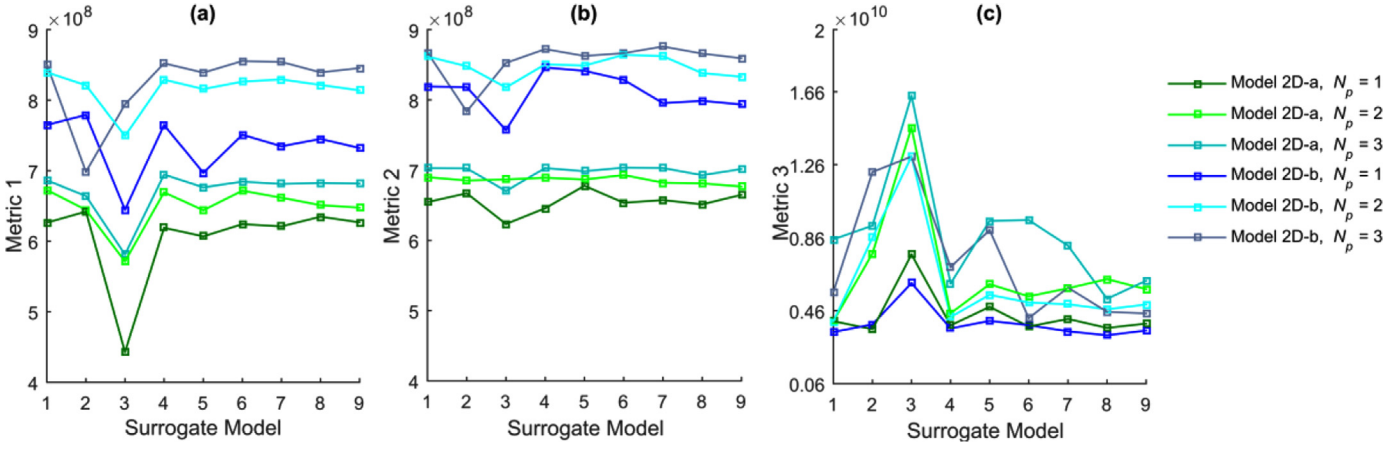


**Fig. 8.** Analysis of Model 2D results: (a) Metric 1 for various surrogate models (refer to Table 1) (b) Metric 2 for various surrogate models, and (c) Metric 3 for various surrogate models.

### 5.2. Results of optimization on the three-dimensional model

Similar to Model 2D, for Model 3D the maximum allowable water injection rate, $q_{inj}^{max}$, is set to 100 surface $m^3$/day. The injection continues for 10 years. We assume two control-period cases, $N_p=1$ and $N_p=2$ to provide 8 and 16 optimization variables. The economical constants are set to $r_o$ = 50 USD/STB, $r_w$ = 10 USD/STB and $r_{inj}$ = 10 USD/STB. We assume no discount factor ($d$ = 0). We set $N_{LHS}$ to $20 \times N_p$ and $N_{eval}$ to $60 \times N_p$. Similar to the two-dimensional models, we ran the optimization problem with each individual or mixed surrogate model 10 times. The results of the optimization are shown in the format of box plots at each step of the optimization in Fig. 9. We show the results for $F(\mathbf{u})$ after the initial design by Latin hypercube sampling only. Also the analyses of results in terms of the metrics defined in the previous subsection are shown in Fig. 10.

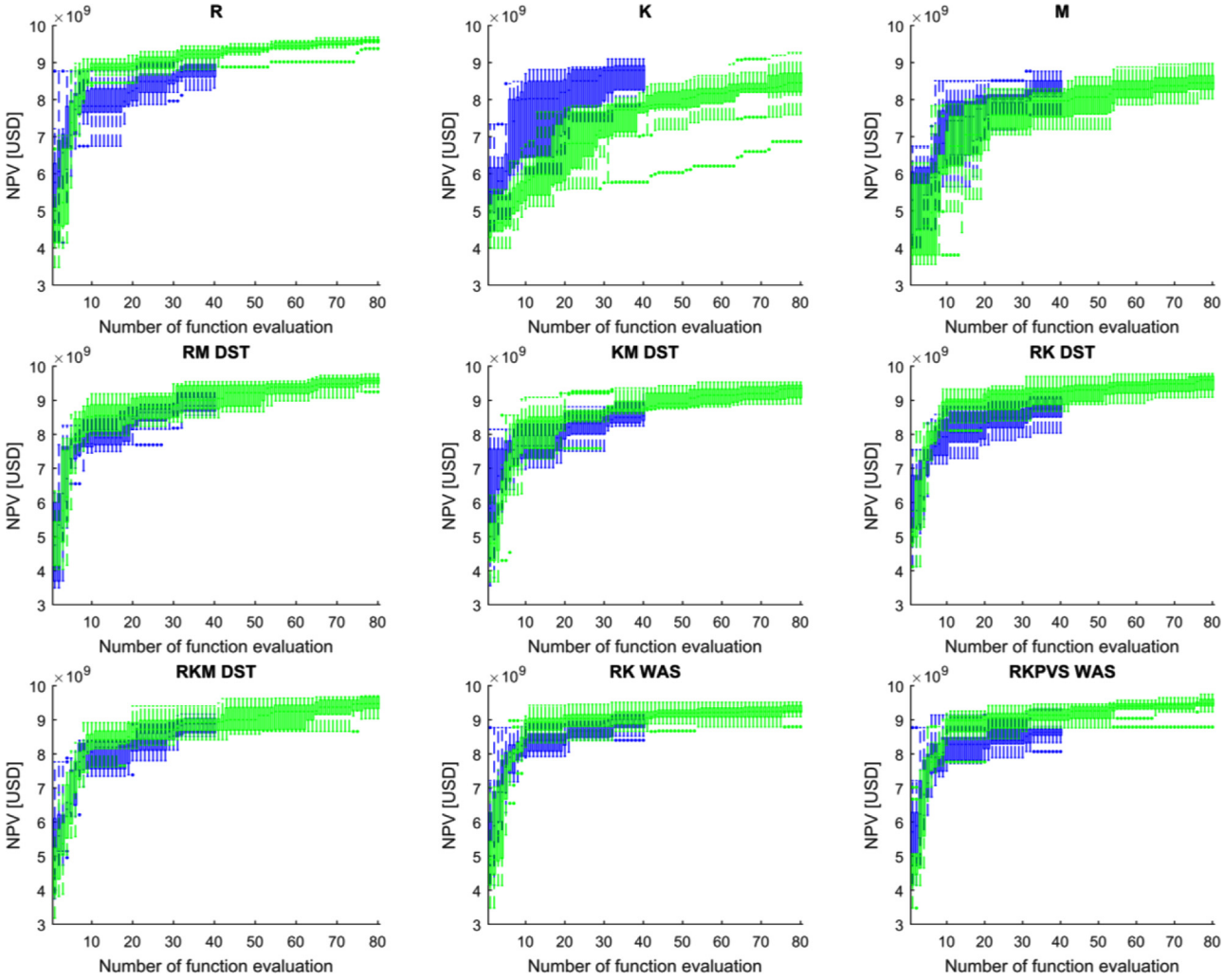


**Fig. 9.** The best solutions obtained per optimization cycles (1 to $N_{eval}$–$N_{LHS}$) by various individual and mixed surrogate models for Model 3D using $N_p=1$ (blue) and Model 3D using $N_p=2$ (green). The number of function evaluations for $N_p=2$ after LHS stage is twice as number of function evaluations for $N_p=1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

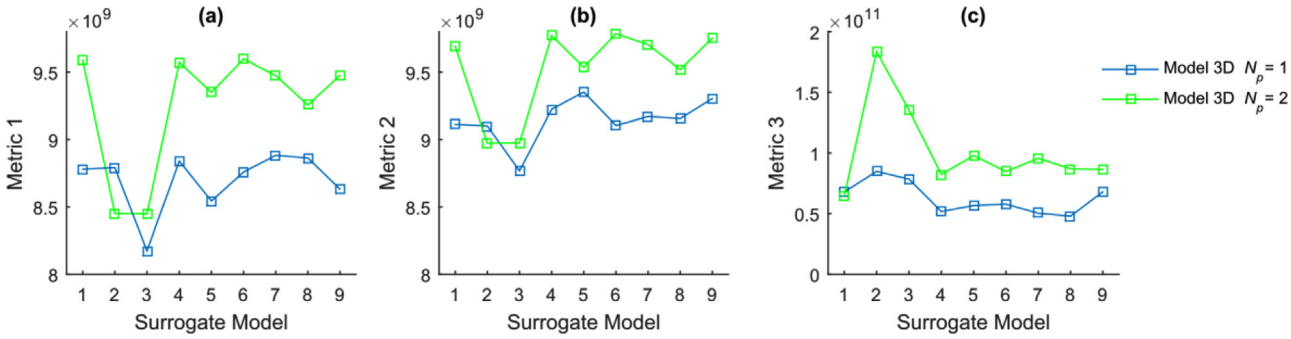


**Fig. 10.** Analysis of Model 3D results: (a) Metric 1 for various surrogate models (refer to Table 1) (b) Metric 2 for various surrogate models, and (c) Metric 3 for various surrogate models.

The analyses of the results in terms of the metrics defined previously, re-affirms the previous conclusions that, firstly, the radial basis function surrogate models perform agreeably well compared to other individual surrogate models and it is not noticeably and significantly outperformed by any of the mixed models in a consistent manner between $N_p=1$ and 2. Interestingly similar to Model 2D, Kriging is performing well for $N_p=1$ in terms of Metrics 1 and 2, but it "catastrophically" fails for $N_p=2$ in terms of all three metrics. This fact demonstrates that Kriging may not be reliable when the numbers of optimization variables are high.

Secondly, between all the possible mixed surrogate models examined, we have found none to be consistently better than radial basis function or better than other mixed surrogate models. It may be a counter–intuitive deduction, however considering the agreeable performance of radial basis function on its own, it may not be too far from expectation that mixing it with other surrogate models by either of the two combination schemes of DST or WAS only undermines its performance.

Lastly, one can easily note the significant improvement of results of Kriging based optimization by implementing mixed surrogate models for $N_p=2$. We believe this was the conclusion of Viana et al. (2013).

In order to illustrate the best solution of optimization, we use the highest *NPV* value obtained by radial basis function, $N_p=2$, where $NPV=9.69\times10^9$ USD. The values of the optimization variables (**u**), for 8 injection wells and two control periods are respectively: [0.22, 0, 0.17, 0.26, 0, 0, 0.03, 0.30, 0, 0.26, 0.03, 0, 0.37, 0.55, 0.38, 0.02]. The injection rate is obtained by multiplying **u** with 100 $m^3$/day. This means that the total water injection rate for the first control period is 98 $m^3$/day and for the second control period is 161 $m^3$/day. The water saturation profiles of the six realizations and the profiles of the injection and production for the wells used in Model 3D are shown in Figs. 11 and 12.

The optimization has found a solution in which the water injection is increased for the second period. This increase coincides with the decrease in oil production rate and increase in water production rate at year 5. Then for a 2 year period, the oil production is maintained high until at year 7 water production increases again. One expects that increasing the number of control periods could potentially increase oil production and subsequently increase net present value of the operation.

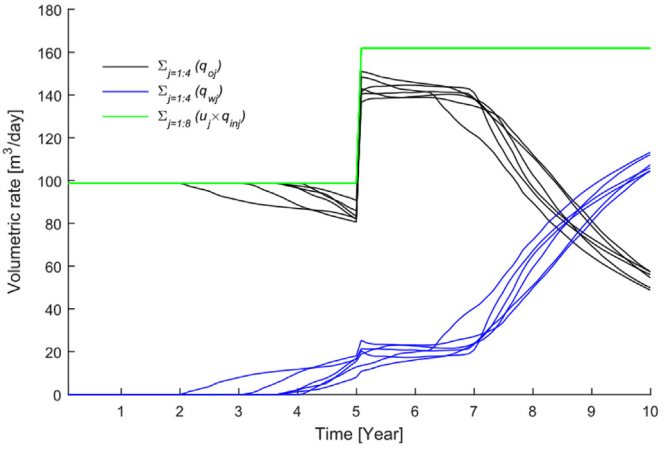


**Fig. 12.** The injection and production curves for 6 realizations used in Model 3D. For each realization, the summation of oil production rate for 4 production wells, water production rate for 4 production wells and water injection rate for 8 injection wells are shown.

## 6. Discussions, conclusions and future works

In this work we compared radial basis function, Kriging and multivariate adaptive regression splines as surrogate models used individually or as ensembles by two available schemes of combination. The optimization problem was defined over two and three dimensional water flooding examples with reasonably high degree of complexity compared to field scale uncertain reservoir engineering models. A derivative-free optimization of expensive black-box objective functions with fixed user-defined parameters was used across all the present models to provide a fair comparison between the models.

Firstly, we did not compare the results of optimization with Genetic Algorithm (GA) as Müller and Shoemaker (2014) have already established the superiorities of surrogate-assisted optimization for the same algorithms. In general GA requires more number of expensive function evaluations for convergence and with a limited budget of function evaluations, it does not give good results. The number of iterations is a proxy for the computational runtimes that are required. This gives a baseline to compare the

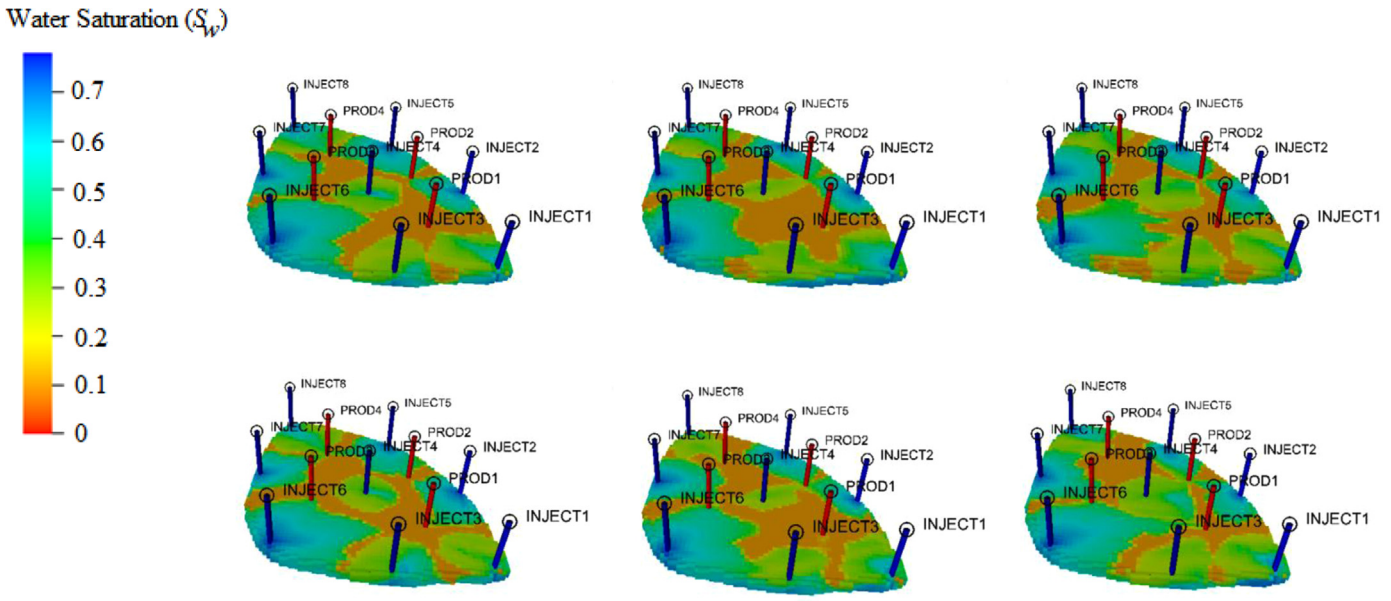


**Fig. 11.** The distribution of water saturation after 10 years of injection for Model 3D and for 6 realizations considered for optimization. This is the results of using the solution of optimization for injection rates obtained from radial basis function surrogate model.

computational efficiency across different hardware platforms and different objective functions as the physical runtime would be different for each of these separate cases. We have done this comparison with GA on similar reservoir engineering problems in (Pan et al., 2014), and it is well-established that the GA does not give good performance with a limited number of function evaluations, in comparison with the surrogate methods.

Secondly, since the optimization algorithm is a stochastic one, it gives slightly different answers every time it is run. To ensure that the results are statistically consistent we ran it multiple times (10 in this case) and used the summary statistics (quantiles, mean, min, max etc.) to compare the accuracy of different variants of the algorithm. This is important since it shows the consistency of the algorithms in finding the best solutions. It might so happen that in one run, one of the algorithms luckily sampled a point very close to the true optima and got a good result, but in general the algorithm is not good at finding such good solutions if it is run multiple times. If our conclusions were based on only one run then they would have been erroneous and therefore there is a need of using multiple runs for comparison.

Our results for individual and ensemble surrogate modeling demonstrate that radial basis function is reliable consistently for different reservoir optimization problems with varying numbers of optimization variables. Even though ensemble surrogate models have been shown to perform somehow better than individual ones on test bench functions (Müller and Piché, 2011; Viana et al., 2013; Müller and Piché, 2014) our results indicate that this is not always the case for realistic reservoir engineering problems. Müller and Piché (2014) showed that the ensemble surrogates outperformed nonsmooth optimization by mesh adaptive direct search (NOMAD) and particle swarm pattern search algorithm (PSWARM), but between the different surrogate models we can hardly choose a consistently superior model. Compared to ensemble models, they showed that for large-dimensional and application problems, R-c (cubic radial basis function with Candidate Point Strategy) has the best average performance. The extensive simulation results on reservoir engineering problems here showed that we may actually jeopardize the performance of cubic radial basis function by combining it with other surrogates.

One plausible reason for this apparent inconsistency in these results might be due to the fact that the objective function of the reservoir engineering problem has complicated underlying nonlinearities which is very different from the ones in the test bench functions. Therefore, even though empirical evidence on the test bench functions have suggested that the DST or WAS based ensemble methods are better than individual surrogates, these do not generalize well to all classes of problems. Therefore these ensemble methods would need to be adapted for similar reservoir engineering problems or new methods of combining multiple surrogates need to be developed, so that they can be useful in such circumstances.

For the case of Viana et al. (2013), we have at least showed that their claim stands true only when the individual model under the comparison is not radial basis function but is others like Kriging for example.

The conclusions of the paper indicate that more work needs to be done on finding an effective surrogate model management strategy which works for realistic reservoir engineering problems. Future research efforts can be geared towards overcoming these shortcomings. Nonetheless it is shown that using radial basis function can be safely used.

In future, we will look at the performance of the algorithms on increasing the number of optimization variables to more accurately reflect a realistic reservoir engineering problem. We would use the methodology for well location optimization for a field-scale realistic reservoir with massively parallelized schemes.

## Acknowledgment

The authors would like to thanks the anonymous reviewers for their invaluable comments.

## Appendix. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.cageo.2016.02.022.

## References

Glaz, B., Goel, T., Liu, L., Friedmann, P.P., Haftka, R.T., 2009. Multiple-surrogate approach to helicopter rotor blade vibration reduction. AIAA J. 47 (1), 271–282. http://dx.doi.org/10.2514/1.40291.

Goel, T., Haftka, R.T., Shyy, W., Queipo, N.V., 2007. Ensemble of surrogates. Struct. Multidiscip. Optim. 33 (3), 199–216. http://dx.doi.org/10.1007/s00158-006-0051-9.

Gorissen, D., Dhaene, T., de Turck, F., 2009. Evolutionary model type selection for global surrogate modelling. J. Mach. Learn. Res. 10, 2039–2078.

Hu, L.Y., Chugunova, T., 2008. Multiple-point geostatistics for modeling subsurface heterogeneity: a comprehensive review. Water Resour. Res. 44 (11). http://dx.doi.org/10.1029/2008WR006993.

Jin, Y., 2005. A comprehensive survey of fitness approximation in evolutionary computation. Soft Comput. 9 (1), 3–12. http://dx.doi.org/10.1007/s00500-003-0328-5.

Jin, Y., 2011. Surrogate-assisted evolutionary computation: Recent advances and future challenges. Swarm Evolut. Comput. 1 (2), 61–70. http://dx.doi.org/10.1016/j.swevo.2011.05.001.

Jin, Y., Olhofer, M., Sendhoff, B., 2002. A framework for evolutionary optimization with approximate fitness functions, Evolutionary Computation. IEEE Trans. Evol. Comput. 6 (5), 481–494. http://dx.doi.org/10.1109/TEVC.2002.800884.

Journel, A.G., 1993. Geostatistics: roadblocks and challenges. In: Geostatistics Tróia'92, Springer, pp. 213–224, doi: 10.1007/978-94-011-1739-5_18.

Lie, K.-A., Krogstad, S., Ligaarden, I.S., Natvig, J.R., Nilsen, H.M., Skaflestad, B., 2012. Open-source Matlab implementation of consistent discretizations on complex grids. Comput. Geosci. 16 (2), 297–322. http://dx.doi.org/10.1007/s10596-011-9244-4.

Müller, J., Piché, R., 2011. Mixture surrogate models based on Dempster-Shafer theory for global optimization problems. J. Glob. Optim. 51 (1), 79–104. http://dx.doi.org/10.1007/s10898-010-9620-y.

Müller, J., Shoemaker, C.A., 2014. Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems. J. Glob. Optim., 1–22. http://dx.doi.org/10.1007/s10898-014-0184-0.

Mariethoz, G., Kelly, B.F.J., 2011. Modeling complex geological structures with elementary training images and transform-invariant distances. Water Resour. Res.. 47 (7). http://dx.doi.org/10.1029/2011WR010412.

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. 46 (11). http://dx.doi.org/10.1029/2008WR007621.

Ong, Y.S., Nair, P.B., Keane, A.J., 2003. Evolutionary optimization of computationally expensive problems via surrogate modelling. AIAA J. 41 (4), 687–696. http://dx.doi.org/10.2514/2.1999.

Pan, I., Babaei, M., Korre, A., Durucan, S., 2014. A multi-period injection strategy based optimisation approach using Kriging meta-models for $CO_2$ storage technologies. Energy Proc. 63, 3492–3499.

Regis, R.G., Shoemaker, C.A., 2007. A stochastic radial basis function method for the global optimization of expensive functions. INFORMS J. Comput. 19 (4), 497–509. http://dx.doi.org/10.1287/ijoc.1060.0182.

Ronayne, M.J., Gorelick, S.M., Caers, J., 2008. Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modelling approach. Water Resour. Res. 44 (8). http://dx.doi.org/10.1029/2007WR006635.

Strebelle, S., 2000. Sequential Simulation Drawing Structures from Training Images (Ph.D. Thesis). Stanford University, CA.

van Essen, G., Zandvliet, M., van den Hof, P., Bosgra, O., Jansen, J.-D., 2009. Robust waterflooding optimization of multiple geological scenarios. SPEJ Soc. Pet. Eng. J. 14 (1), 202–210. http://dx.doi.org/10.2118/102913-PA.

Viana, F.A.C., Haftka, R.T., Steffen Jr, V., 2009. Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Struct. Multidiscip. Optim. 39 (4), 439–457.

Viana, F.A.C., Haftka, R.T., Watson, L.T., 2013. Efficient global optimization algorithm assisted by multiple surrogate techniques. J. Glob. Opt. 56 (2), 669–689.

Viana, F.A.C., Haftka, R.T., 2008. Using multiple surrogates for metamodeling. Paper Presented at the Seventh ASMO-UK/ISSMO International Conference on Engineering Design Optimization, Bath, UK, July 7– 8.

Wu, J., Boucher, A., Zhang, T., 2008. A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. Comput. Geosci. 34 (12), 1863–1876. http://dx.doi.org/10.1016/j.cageo.2007.08.008.

Yoon, J.-H., Shoemaker, C.A., 1999. Comparison of optimization methods for ground-water bioremediation. J. Water Resour. Plan. Manag. 125 (1), 54–63. http://dx.doi.org/10.1061/(ASCE)0733-9496(1999)125:1(54)).

Zhang, J., Chowdhury, S., Messac, A., 2012. An adaptive hybrid surrogate model. Struct. Multidiscip. Optim. 46 (2), 223–238. http://dx.doi.org/10.1007/s00158-012-0764-x.

Zhang, T., Switzer, P., Journel, A., 2006. Filter-based classification of training image patterns for spatial simulation. Math. Geol. 38 (1), 63–80. http://dx.doi.org/10.1007/s11004-005-9004-x.

Zhou, Z., Ong, Y.S., Lim, M.H., Lee, B.S., 2007. Memetic algorithm using multi-surrogates for computationally expensive optimization problems. Soft Comput. 11 (10), 957–971. http://dx.doi.org/10.1007/s00500-006-0145-8.

Zubarev, D.I., 2009. Pros and cons of applying proxy-models as a substitute for full reservoir simulations. Paper Presented at SPE Annual Technical Conference and Exhibition, 4–7 October, New Orleans, LA, doi: http://dx.doi.org/10.2118/124815-MS.