

Reservoir Geological Uncertainty Reduction: an Optimization-Based Method Using Multiple Static Measures

Shahed Rahim · Zukui Li · Japan Trivedi

Received: 23 March 2014 / Accepted: 29 November 2014 / Published online: 14 January 2015
© International Association for Mathematical Geosciences 2014

Abstract Uncertainty in reservoir geological properties has a major impact in reservoir design and operations decision-making. To quantify the production uncertainty and to make optimal decisions in reservoir development, flow simulation is widely used. However, reservoir flow simulation is a computationally intensive task due to complex geological heterogeneities and numerical thermal modeling. Normally only a small number of realizations are chosen from a large superset for flow simulation. In this paper, a mixed-integer linear optimization-based geological realization reduction method is proposed to select geological realizations. The method minimizes the probability distance between the discrete distribution represented by the superset of realizations and the reduced discrete distribution represented by the selected realizations. The proposed method was compared with the traditional ranking technique and the distance-based kernel clustering method. Results show that the proposed method can effectively select realizations and assign probabilities such that the extreme and expected reservoir performances are recovered better than any of the single static measure-based ranking methods or the kernel clustering method.

Keywords Reservoir uncertainty · Optimal realization reduction · Mixed-integer optimization · Static measures

S. Rahim · Z. Li (✉)
Department of Chemical and Materials Engineering,
University of Alberta, Edmonton, Alberta, Canada
e-mail: zukui@ualberta.ca

J. Trivedi
Department of Civil and Environmental Engineering,
University of Alberta, Edmonton, Alberta, Canada

1 Introduction

Reservoir geological properties are important parameters used in the design and optimization of oil extraction processes from reservoirs. These parameters dictate the ease with which oil can be extracted and also the quantity of oil that can be extracted. Geological uncertainty exists because it is not possible to know the exact geological properties of every section of a realistic reservoir. Techniques such as well exploration and core holes can give an idea of the geological properties of particular areas of the reservoir. However, the geological parameters of the areas between the exploration wells or core holes will still be unknown. As a result, geological uncertainty will always exist for a reservoir.

Reservoir performance can be quantified by flow simulation which provides production parameters of interest such as the cumulative oil production (COP) rate and the net present value (NPV). All of the production parameters depend on the geological properties of the reservoir. It is very important to incorporate geological uncertainty in a reservoir model. Otherwise, the model may give an incorrect prediction of production parameters. To represent the geological uncertainty, multiple geological realizations are usually generated using geostatistical tools so as to obtain a broad range of possible geological properties for a reservoir. However, reservoir flow simulations cannot be run for all of the possible realizations due to the significant computer processing time. Therefore, in practice, only a small number of geological realizations are chosen to perform reservoir simulations to obtain a reservoir performance model which incorporates geological uncertainty. Various methods for selecting geological realizations exist in the literature and can be broadly classified as follows: random selection method, static measure-based ranking method, distance-based kernel clustering technique and probability distance-based realization reduction method.

While randomly selecting a subset of realizations is the easiest method for implementation, it may result in the wrong measure of geological uncertainty especially when the number of selected realizations is small. Many studies in the literature use the single static measure-based ranking method to select geological realizations and to quantify the uncertainty in reservoir performance. The ranking method was introduced by [Ballin et al. \(1992\)](#). Ranking-based reduction arranges realizations of an easily computable measure in an ascending/descending order and then selects the realizations that have low, medium and high measure values. The selected realizations are then used as input for flow simulations to obtain the reservoir production response. [Deutsch \(1998, 1999\)](#) developed software tools to rank realizations based on the number of connected cells, connectivity to a well location or connectivity between multiple wells. [Deutsch and Begg \(2001\)](#) proposed that ranking all of the realizations based on a measure and then choosing a subset of equally spaced realizations result in a better representation of uncertainty than choosing the low-, medium- and high-performance realizations. A ranking-based method for the Steam-Assisted Gravity Drainage (SAGD) process using a measure known as connected contained bitumen was used by [McLennan and Deutsch \(2004\)](#). The connected contained bitumen was calculated using net cells connected to the SAGD producer well. The realizations with low, medium and high ranking measures were selected from the superset of realizations. In another study, [McLennan and Deutsch \(2005\)](#) used several measures based on statistical, volumet-

ric, global and local connectivity metrics to select a subset of realizations. [Fenik et al. \(2009\)](#) used a ranking method based on connected hydrocarbon volume (CHV) to select a subset of realizations for a SAGD application. [Li et al. \(2012\)](#) adopted a static quality measure, which was a modification to the CHV measure, to rank geological realizations. Rankings based on the static quality measure showed improved performance over rankings based on CHV. While efforts have been made to improve static measures for ranking, the limitation of existing ranking methods for selecting realizations is that they rely greatly on the measure used. If the measure has poor correlation to the production performance parameter of the reservoir, then the selected realizations will not be a good representation of the superset of realizations. Furthermore, all of the selected realizations based on the ranking method have equal probability in the reduced distribution.

The distance-based kernel clustering method has also received lots of attention in the past. [Scheidt and Caers \(2009\)](#) used simplified streamline simulation results to compute the distance between realizations and to form a distance matrix. The uncertainty associated with the distance matrix is modelled using multidimensional scaling and kernel techniques. The superset of realizations is grouped into clusters using kernel k-means clustering, and a subset of selected realizations can be extracted from the clusters. [Scheidt and Caers \(2010\)](#) compared the statistics associated with the traditional ranking method, the kernel k-means clustering method and the random selection method. They used the bootstrap technique to compute the confidence intervals of the P10, P50 and P90 quantiles of the reduced subset of realizations and showed that the distance-based kernel clustering method provides the most robust results ([Scheidt and Caers 2010](#); [Park and Caers 2011](#)). [Singh et al. \(2014\)](#) used the kernel k-means clustering method to quantify uncertainty associated with various history-matched geological models and to forecast production information. The distance matrix in the clustering method uses oil recovery factors between realizations. These clustering methods generally rely on streamline simulations to calculate distance between realizations, which is still computationally demanding for large reservoir models, and the problem has to be reformulated if the number of wells or well location changes. Additionally, simplified fluid flow assumptions are used for streamline simulations, which may undermine the geological heterogeneity of the reservoir, resulting in a poor representation of geological uncertainty ([Gilman et al. \(2002\)](#)). There is a need for a realization reduction method that is computationally less expensive and provides a better representation of the original distribution than the current ranking methods.

Apart from the ranking method and the clustering method, a probability distance-based realization reduction/selection method associated with a scenario reduction technique for optimization under uncertainty [[Dupacova et al. 2003](#); [Li and Floudas 2014](#)] has been investigated recently. Following this direction, [Armstrong et al. \(2013\)](#) proposed a realization reduction method based on minimizing the Kantorovich distance between distributions and applied the method to metal mining. The method iteratively generates a random subset of realizations without replacements from the superset of realizations until the Kantorovich distance between the distributions is minimized. While their method relies on heuristic random searches to minimize the Kantorovich distance, a novel method is proposed in this work for geological realization reduction following the concept of probability distance minimization. The pro-

posed method considers multiple static measures and geological properties to select geological realizations. Specifically, an optimal realization reduction model is developed based on the mixed-integer linear optimization (MILP) technique. The proposed algorithm uses reservoir geological properties and static measures to quantify the dissimilarity between realizations and uses the Kantorovich distance to quantify the probability distance between the superset and the subset of realizations. The objective is to find the optimal subset that has a similar statistical distribution characteristic to the superset of realizations.

The remainder of the paper is organized as follows: Sect. 2 provides an overview of different static measures used in this study; Sect. 3 provides a detailed description of the proposed optimal realization reduction algorithm; Sect. 4 provides a reservoir case study and the results; and conclusions are presented in Sect. 5.

2 Static Measures

Static measures are simplified metrics designed to achieve a good correlation with the reservoir production performance variable of interest. Compared to flow simulation, static measure is computationally much easier for evaluation. It can be easily computed for a large set of realizations. Static measures can be classified into the following categories [Deutsch and Srinivasan 1996; McLennan and Deutsch 2005]: (i) statistical static measures which quantify the statistical average of geological parameters, (ii) fractional static measures that determine the active fraction of the reservoir, and (iii) volumetric static measures which calculate the volume of a reservoir capable of oil transport. Details on the different static measures are given in the following subsections.

2.1 Statistical Static Measures

Statistical static measures considered in this paper are calculated for net cells in the reservoir. Any cell which has a porosity and permeability above a threshold value is defined as a net cell. A binary indicator parameter I_c^{net} is used to denote whether a cell c in the reservoir grid is net ($I_c^{\text{net}} = 1$) or not ($I_c^{\text{net}} = 0$). The idea of a net cell stems from the fact that if a section of the reservoir rock has very low porosity and permeability value, then that section of the rock will be unable to carry any oil through it. As a result, that non-net section of the rock plays no role in oil recovery from the reservoirs. Mathematically I_c^{net} is defined as

$$I_c^{\text{net}} = \begin{cases} 1, & \text{if } \phi_c \geq \phi_0 \text{ and } k_c \geq k_0 \\ 0, & \text{otherwise} \end{cases} . \quad (1)$$

In Eq. (1), ϕ_c and k_c denote the porosity and the permeability of cell c , respectively, whereas ϕ_0 and k_0 denote the threshold values.

Statistical static measures are the simplest measures for ranking realizations. The average net permeability (K_{net}) for each realization is given by Eq. (2)

$$K_{\text{net}} = \frac{\sum_c k_c I_c^{\text{net}}}{\sum_c I_c^{\text{net}}}. \quad (2)$$

Indicator I_c^{net} is used in Eq. (2) to ensure that the average permeability is only calculated for the net cells. Using a similar idea, the average net porosity (ϕ_{net}) for each realization is given by Eq. (3)

$$\phi_{\text{net}} = \frac{\sum_c \phi_c I_c^{\text{net}}}{\sum_c I_c^{\text{net}}}. \quad (3)$$

The average net irreducible water saturation (S_{net}) for each realization is given by Eq. (4)

$$S_{\text{net}} = \frac{\sum_c S_c I_c^{\text{net}}}{\sum_c I_c^{\text{net}}}, \quad (4)$$

where S_c is the irreducible water saturation of cell c . Similar to average permeability and porosity, average irreducible water saturation is only calculated for the net cells.

2.2 Fractional Static Measures

Fractional static measures use indicator parameters to calculate the fraction of cells that are either net or locally connected to a well. These static measures provide a good basis for understanding the quality of a reservoir and the amount of oil that can be extracted from a reservoir.

The fraction of net cells (F_{net}) of a reservoir is also known as the net to gross ratio. It is calculated by the summation of I_c^{net} values of each cell and divided by the total number of cells N as given in Eq. (5) below

$$F_{\text{net}} = \frac{1}{N} \sum_c I_c^{\text{net}}. \quad (5)$$

As the number of net cells in a reservoir increases, the net to gross ratio increases as well. Therefore, a higher net to gross ratio implies that the reservoir will have better oil production.

The fraction of locally connected cells (F_{LC}) is the fraction of cells that are net and are connected to a producer well. A cell c is defined as locally connected if $I_c^{\text{net}} = 1$ and there is a path of net cells from cell c to a producer well. Therefore, all locally connected cells are net cells, but net cells are not necessarily locally connected cells. Mathematically I_c^{LC} is denoted by Eq. (6)

$$I_c^{\text{LC}} = \begin{cases} 1, & \text{if } I_c^{\text{net}} = 1 \text{ and connected to producer well} \\ 0, & \text{if } I_c^{\text{net}} = 0 \text{ or not connected to producer well.} \end{cases} \quad (6)$$

The fraction of locally connected cells is calculated using the binary variables I_c^{net} and I_c^{LC} as given by Eq. (7)

$$F_{\text{LC}} = \frac{1}{N} \sum_c I_c^{\text{net}} I_c^{\text{LC}}. \quad (7)$$

Since the local connectivity calculation considers only active cells connected to a well, it is shown to be a good indication of production parameters such as COP or NPV.

2.3 Volumetric Static Measures

Volumetric static measures incorporate the volume of each cell in its calculation and, therefore, provide a good basis for determining the volume of oil each cell in the reservoir can produce.

Net pore volume (PV_{net}) is the simplest volumetric static measure that utilizes the volume of each cell and the corresponding porosity of that cell. The calculation of net pore volume is only for net cells since these are the only cells that have the ability to produce or transport oil. The net pore volume is given by Eq. (8) below

$$PV_{\text{net}} = \sum_c V_c \phi_c I_c^{\text{net}}, \quad (8)$$

where V_c is the volume of cell c .

Original oil-in-place (OOIP) is calculated for all cells in the reservoir and is calculated by the summation of the product of volume (V_c), porosity (ϕ_c) and the oil saturation ($1 - S_c$) of cell c . OOIP is given by Eq. (9) below

$$\text{OOIP} = \sum_c V_c \phi_c (1 - S_c), \quad (9)$$

where S_c is the irreducible water saturation of cell c .

Net oil-in-place (OOIP_{net}), which is also known as net hydrocarbon volume, is the OOIP for the net cells of the reservoir. Therefore, in principle, OOIP_{net} is expected to be a better static measure than OOIP. OOIP_{net} is given by Eq. (10) below

$$\text{OOIP}_{\text{net}} = \sum_c V_c \phi_c (1 - S_c) I_c^{\text{net}}. \quad (10)$$

Locally connected hydrocarbon volume ($\text{CHV}_{\text{local}}$) is the OOIP calculated for net cells connected to the producer well (Deutsch 1998). A cell is considered to be locally connected when an active cell pathway can be formed from the cell to a producer well so that the oil can be transported (McLennan and Deutsch 2004). Locally connected hydrocarbon volume is given by Eq. (11)

$$\text{CHV}_{\text{local}} = \sum_c V_c \phi_c (1 - S_c) I_c^{\text{net}} I_c^{\text{LC}}. \quad (11)$$

3 Proposed Realization Reduction Algorithm

3.1 Dissimilarity Between Realizations

Considering two geological realizations i and i' , a dissimilarity measure is used to quantify the difference between them. In this work, the dissimilarity between realizations is computed using the geological properties and the static measures introduced in the previous section. Specifically, the dissimilarity between two realizations i and i' is evaluated by Eq. (12)

$$c_{i,i'} = \sum_k |m_{ik} - m_{i'k}| + \sum_{c,t} \gamma |\theta_{ict} - \theta_{i'ct}| \quad \forall i, i', \quad (12)$$

where m_{ik} is the value of the k -type static measure for realization i , θ_{ict} is the t -type geological property value of cell c in the reservoir grid for realization i and γ is a weight parameter which reflects the contribution of geological property data in the dissimilarity calculation. The static measure parameters are given a larger weight here to emphasize its importance in the dissimilarity evaluation (γ is set as 0.01 in this work). Since static measures are easily computable for any given realization, the dissimilarity values between any two geological realizations can be calculated very efficiently.

3.2 Kantorovich Distance in Realization Reduction

All of the geological realizations generated from geostatistical tools form a superset from which a subset containing a small number of realizations is to be selected for further investigation (e.g., flow simulation). The objective of realization reduction is that the selected subset of realizations can represent the superset of realizations very well in terms of reservoir production performance.

In this work, the aforementioned superset and subset are considered as two discrete probability distributions. The first distribution (also called the original distribution) consists of the superset of realizations, and each realization i has probability p_i^{orig} . Notice that this probability value p_i^{orig} is normally set as equal to $1/|I|$, where I is the superset of realizations and $|I|$ is the set cardinality (i.e., total number of realizations). The second distribution (also called the reduced distribution) consists of a subset of selected realizations in which each realization i has probability p_i^{new} . The reduced distribution can be viewed as an updated version of the original distribution with probabilities on each realization adjusted. All of the removed realizations have zero probability in the reduced distribution.

The Kantorovich distance is a type of probability metric to quantify the dissimilarity between two probability distributions. It is defined by a transportation problem which minimizes the transportation cost associated with moving the probability mass from one distribution to the other distribution. The theory of optimal transportation was first introduced by Monge (1781) and rediscovered by Kantorovich (1942). For the realization reduction problem, the Kantorovich distance between the original dis-

tribution and the reduced distribution is defined by the optimal objective value of the following linear transportation problem

$$\begin{aligned}
 D_{\text{Kan}} &= \min_{\eta_{i,i'}} \sum_{i \in I} \sum_{i' \in S} \eta_{i,i'} c_{i,i'} \\
 \text{s.t. } \sum_{i \in I} \eta_{i,i'} &= p_{i'}^{\text{new}} \quad \forall i' \in S \\
 \sum_{i' \in S} \eta_{i,i'} &= p_i^{\text{orig}} \quad \forall i \in I \\
 \eta_{i,i'} &\geq 0 \quad \forall i \in I \quad \forall i' \in S,
 \end{aligned} \tag{13}$$

where i and i' represent realizations, I is the superset and S is the selected subset, p_i^{orig} and p_i^{new} represent the probability of realization i in the original and the reduced distribution, respectively, $\eta_{i,i'}$ is the decision variables representing the probability mass transportation plan and $c_{i,i'}$ is the dissimilarity between realizations. Dupacova et al. (2003) proved that the optimal objective value of the above problem is

$$D_{\text{Kan}} = \sum_{i \in I-S} p_i^{\text{orig}} d_i, \tag{14}$$

where $d_i = \min_{i' \in S} c_{i,i'}$ represents the transportation cost for a removed realization $i \in I - S$ and it is the minimum dissimilarity between the removed realization i and all of the selected realizations $i' \in S$. The optimal solution of $p_{i'}^{\text{new}}$ for problem (13) is

$$p_{i'}^{\text{new}} = p_{i'}^{\text{orig}} + \sum_{i \in J(i')} p_i^{\text{orig}} \quad \forall i' \in S, \tag{15}$$

where $J(i') = \{i | i \in I - S, c_{i,i'} \leq c_{i,i''}, \forall i'' \in S\}$, meaning that a preserved realization's new probability is the sum of its original probability and the probability mass that has been transported to it. A removed realization is transported to the closest preserved realization.

3.3 Realization Reduction Algorithm

To select representative geological realizations, an optimization-based realization reduction method is proposed in this work. The proposed realization selection/reduction method is based on a constrained mixed-integer linear optimization technique, and it minimizes the Kantorovich distance between the original distribution and the reduced distribution as explained in previous subsection. Details on the proposed optimization model are stated below.

First, binary variables y_i are introduced to denote whether the realization i is removed ($y_i = 1$) or not ($y_i = 0$). Continuous variables $v_{i,i'}$ ($0 \leq v_{i,i'} \leq 1$) are

introduced to denote the fraction of the probability mass that is transported from realization i to realization i' .

The objective function of the proposed realization reduction algorithm is to minimize the Kantorovich distance between the original distribution and the reduced distribution, which is given in Eq. (16)

$$\min D_{\text{Kan}} = \sum_{i \in I} p_i^{\text{orig}} d_i, \tag{16}$$

where d_i represents the cost of removing a realization i (i.e., transporting and distributing its probability mass to preserved realizations). This cost is quantified by a weighted summation of the transported probability mass, where the weight is the dissimilarity $c_{i,i'}$ between realizations. This scheme can be modeled using Eq. (17)

$$d_i = \sum_{i' \in I} c_{i,i'} v_{i,i'} \quad \forall i \in I. \tag{17}$$

Notice that the Kantorovich distance defined in problem (13) is based on a known subset S , while the objective in the proposed algorithm here is to find the optimal subset S that leads to the minimum Kantorovich distance. With the introduction of variables y_i and $v_{i,i'}$, the proposed optimization model will generate the optimal S that leads to the minimum Kantorovich distance as explained by the following constraints. The reader is also referred to Li and Floudas (2014) for detailed proof on this.

The following set of constraints are necessary to enforce the logical relationship between variables y_i and $v_{i,i'}$. First, if a realization i is removed ($y_i = 1$), then all of its probability mass should be transported ($\sum_{i' \in I} v_{i,i'} = 1$). If a realization i is selected/preserved ($y_i = 0$), then its probability mass should not be transported to any realization ($\sum_{i' \in I} v_{i,i'} = 0$). The above logical relationship is reflected in Eq. (18)

$$\sum_{i' \in I} v_{i,i'} = y_i \quad \forall i \in I. \tag{18}$$

Furthermore, if a realization i' is removed ($y_{i'} = 1$), then no probability mass can be transported to it ($v_{i,i'} = 0$). If a realization i' is selected ($y_{i'} = 0$), then the probability mass can be transported to it ($0 \leq v_{i,i'} \leq 1$). This condition is modeled in Eq. (19) below

$$0 \leq v_{i,i'} \leq 1 - y_{i'} \quad \forall i, i' \in I. \tag{19}$$

The next constraint enforces the number of selected realizations. Assume the total number of realizations to be removed is R , then Eq. (20) ensures that R realizations are removed

$$\sum_{i \in I} y_i = R. \tag{20}$$

In the proposed realization reduction model, a subset of realizations representing the potential best and worst performance is also considered. Equation (21) below ensures that at least two realizations are selected from subset I_{SB}

$$\sum_{i \in I_{SB}} (1 - y_i) \geq 2, \quad (21)$$

where subset I_{SB} has two realizations which are identified using the following steps. For each static measure, the realizations corresponding to the top three highest static measure values are identified. Those identified realizations' IDs are combined into a superset from which the two most frequent realizations are selected to form set I_{SB} . Similarly, Eq. (22) ensures that at least two realizations are selected from subset I_{SW} in the reduced distribution

$$\sum_{i \in I_{SW}} (1 - y_i) \geq 2, \quad (22)$$

where subset I_{SW} has the top two most frequent realizations that represent the potential worst performance. For each static measure, the realizations corresponding to the top three lowest static measure values are identified. Those identified realizations' IDs are combined into a superset from which the two most frequent realizations are selected to form set I_{SW} .

With the selected realizations (i.e., y_i) and the probability mass transportation plan (i.e., $v_{i,i'}$), the new probability of realizations in the reduced distribution p_i^{new} can be evaluated as follows

$$p_{i'}^{\text{new}} = (1 - y_{i'})p_{i'}^{\text{orig}} + \sum_i v_{i,i'}p_i^{\text{orig}} \quad \forall i' \in I. \quad (23)$$

Notice that if realization i' is removed ($y_{i'} = 1$), then $p_{i'}^{\text{new}} = 0$. If realization i' is preserved ($y_{i'} = 0$), then its new probability mass can be calculated as the sum of all of the probability mass that has been transported to it ($\sum_i v_{i,i'}p_i^{\text{orig}}$) and its original probability ($p_{i'}^{\text{orig}}$).

Finally, the complete optimization model is composed of Eqs. (16)–(23), and it is a MILP optimization problem. This problem can be solved using a MILP solver such as CPLEX (IBM 2010). The complete model is given in Appendix B.

4 Case Study

To study the proposed geological realization reduction method, a three-dimensional reservoir model with a grid size of $60 \times 220 \times 5$ (66,000 total cells) and cell sizes of $6.096 \text{ m} \times 3.048 \text{ m} \times 0.6096 \text{ m}$ is investigated in this section. The reservoir has one vertical injector well placed at grid position [30 110] and four vertical producer wells placed at [1 1], [60 1], [60 220] and [1 220]. The reservoir grid size, fluid properties and well locations for the case study were obtained from the SPE 10 comparative solution

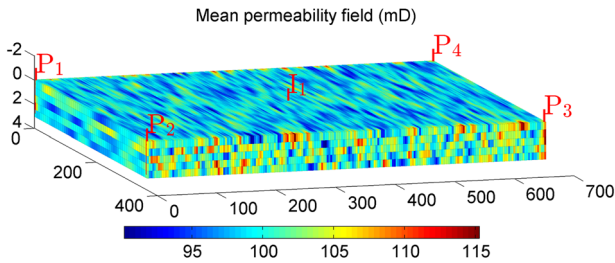


Fig. 1 Three-dimensional grid structure of the mean permeability distribution of the reservoir with the injector and producer locations

Table 1 Case study parameters

Parameter	Value
Initial pressure p_i	6,000 psi
Oil viscosity μ_o at p_i	3 cp
Water viscosity μ_w at p_i	0.3 cp
Oil density ρ_o	849 kg/m ³
Water density ρ_w	1,025 kg/m ³
Relative permeability exponent for oil n_o	2
Relative permeability exponent for water n_w	2
Residual phase saturation for oil S_{r_o}	0.2
Residual phase saturation for water S_{r_w}	0.2
Relative permeability for oil k_{w_o} at S_{r_o}	1
Relative permeability for water k_{w_w} at S_{r_w}	1
Oil Price	100\$/stb
Water production cost	10\$/stb
Water injection cost	10\$/stb
Discount Rate	0 %

project model 2 (Christie and Blunt 2001). Figure 1 shows the three-dimensional grid with the mean permeability field from the superset of realizations and the locations of the vertical injector well (I_1) and the four producer wells (P_1 , P_2 , P_3 , P_4). Detailed fluid properties and cost data used for the simulation are provided in Table 1.

To check the performance of the realization reduction, the Matlab Reservoir Simulation Toolbox (MRST) (Lie et al. 2012) was used to perform reservoir simulations on different geological realizations to obtain the production parameters for validation. The production parameters evaluated in the simulator include NPV, COP and water cut. The simulation time horizon is set as 360 days and is divided into 12 equal periods.

In this study, 100 realizations were generated for realization reduction. For each realization, porosity values of the reservoir grid were generated in MRST using a built-in function ‘Gaussian Field’ with a range parameter of [0.2 0.4]. The function creates an approximate Gaussian random field by convolving a normal distributed random field with a Gaussian filter with a standard deviation of 2.5 (Lie et al. 2012). For illustrative purposes, the porosity distribution of the seventy-fifth realization is

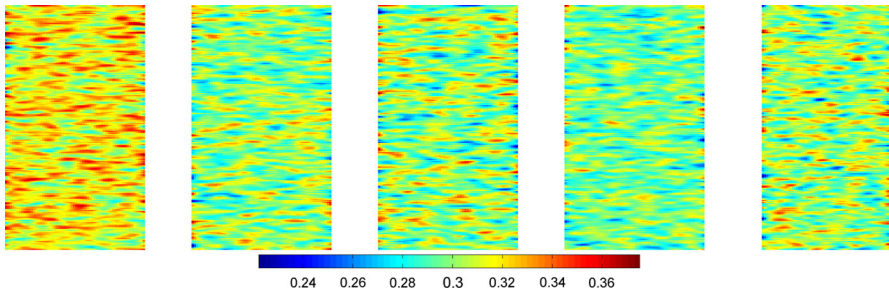


Fig. 2 Porosity distribution of the grid layers for the seventy-fifth realization from the superset

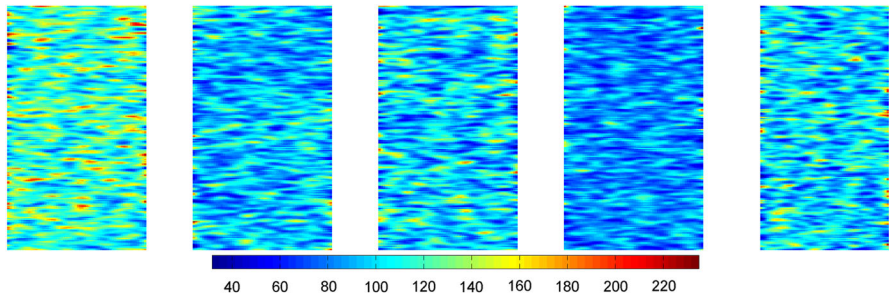


Fig. 3 Permeability (mD) distribution of the grid layers for the seventy-fifth realization from the superset

given in Fig. 2. Each subfigure represents a layer of the three-dimensional grid. From left to right, the subfigures represent the porosity distribution of layers 1 to 5 of the grid.

Permeability values were further generated from the porosity values using the Carman–Kozeny relationship (Lie et al. 2012)

$$k_c = \frac{1}{2\tau A_v^2} \frac{\phi_c^3}{(1 - \phi_c)^2}. \tag{24}$$

In Eq. (24), k_c is the permeability of cell c , ϕ_c is the porosity of cell c , A_v is the surface area of spherical uniform grains with a constant diameter of 10×10^{-6} m and τ is the tortuosity with a value of 0.81 (Lie et al. 2012). The irreducible water saturation (S_c) values for each cell c were generated next from the porosity and permeability values using the Wyllie and Rose equation as given in Eq. (25)

$$k_c = \left[\frac{100\phi_c^{2.25}}{S_c} \right]^2. \tag{25}$$

The permeability distribution of the seventy-fifth realization based on the above calculation is given in Fig. 3.

To evaluate the static measures for different geological realizations in this case study, the threshold porosity is set as $\phi_0 = 0.25$ and the threshold permeability is

Table 2 Realizations selected using the proposed method

Realization ID	Probability	NPV (\$)	COP (m ³ /day)
3	0.04	43,207,700	114,482
8	0.2	50,161,200	124,532
20	0.11	46,223,000	118,840
41	0.14	51,491,500	126,455
64	0.01	55,497,500	132,245
76	0.15	47,146,400	120,175
79	0.05	53,198,600	128,922
80	0.01	43,664,800	115,143
85	0.19	49,291,600	123,274
99	0.1	48,695,600	122,414

set as $k_0 = 1 \times 10^{-13} \text{ m}^2$ to determine whether a cell is net or not. In this case study, the realization reduction results from the proposed method are compared to the traditional ranking method, kernel k-means clustering method and random selection method. A subset of 10 realizations was selected from the superset of 100 realizations to investigate the effectiveness of the proposed method.

4.1 Proposed Method

Based on the generated geological realizations, different static measures stated in Sect. 2 were calculated first. Those measures were further used to compute the dissimilarities between realizations as defined in Eq. (12). The proposed MILP model was built next, which has 100 binary variables corresponding to each realization in the superset. The optimization problem was solved using the CPLEX solver in a desktop computer with a 3.2 GHz processor and 8 GB memory in less than 1 s. The solution of the optimization problem includes the selected realizations' IDs and their new probabilities. For comparison purposes, the selected realizations' IDs, probabilities, NPV and COP values associated with the selected realizations are reported in Table 2.

4.2 Ranking Method

The static measure-based ranking method was applied next to obtain a subset of 10 selected realizations from the superset of 100 realizations. In the ranking-based methods, all of the 100 realizations in the superset were sorted in ascending order based on the static measure values. Ten realizations were evenly selected from the sorted list corresponding to ranks of 1, 12, 23, 34, 45, 56, 67, 78, 89 and 100. In this study, nine different static measures were used to perform realization reduction using the ranking-based method.

Results from the ranking method are just the realizations' IDs. The probabilities of the selected realizations are assumed to be equal (i.e., 0.1 in this case). The selected

Table 3 Realizations selected using kernel k-means clustering

Realization ID	NPV (\$)	COP (m ³ /day)
19	49,887,900	124,137
99	48,695,600	122,414
78	51,281,700	126,152
66	50,017,000	124,324
70	45,745,300	118,150
92	53,209,000	128,937
30	50,347,000	124,801
87	51,628,000	126,652
37	49,306,000	123,296
24	47,640,100	120,888

realizations' IDs, the corresponding static measure values, COP and NPV value from flow simulation are reported in Table 6 of Appendix A.

4.3 Kernel k-Means Clustering Method

The kernel k-means clustering method proposed by [Scheidt and Caers \(2010\)](#) was applied to this case study. The workflow is as follows:

1. Computation of the dissimilarity matrix. While the literature generally uses simplified flow simulation results to calculate the dissimilarity, the same dissimilarity values used by the proposed method were used for clustering in this study.
2. Classical Multidimensional Scaling (MDS) was used to transform the dissimilarity matrix into reduced dimensional data in Euclidean space.
3. Conversion of the Euclidean space to a feature space using a Gaussian kernel (radial basis function) given by Eq. (26)

$$K_{mn} = \kappa(x_m, x_n) = \exp\left[-\frac{\|x_m - x_n\|}{2\sigma^2}\right]. \quad (26)$$

In the Gaussian kernel function, a kernel width, σ , value between 10 and 20 % of the range of the distance measures was used based on the recommendation by [Shi and Malik \(2000\)](#). $\sigma = 2,000$ was used in this study.

4. Ten clusters were generated using k-means clustering. The realization ID closest to the centroid of each cluster was chosen as the representative realization from the kernel k-means clustering.

The realization ID, NPV and COP values from the 10 selected realizations using kernel k-means clustering are given in Table 3. All of the selected realizations of the clustering method have equal probabilities.

Table 4 Realizations selected using random selection

Realization ID	NPV (\$)	COP (m ³ /day)
12	52,791,900	128,334
15	49,003,702	122,859
44	49,679,300	123,836
36	51,044,700	125,809
51	50,191,100	124,575
2	49,371,500	123,391
81	49,966,000	124,250
61	52,327,100	127,663
75	48,834,300	122,614
67	48,086,500	121,534

4.4 Random Selection

A subset of realizations was selected randomly to compare with the proposed method. Random permutation was used to generate 10 realization IDs from the superset. The realization ID, NPV and COP values from 10 equiprobable randomly selected realizations are given in Table 4.

4.5 Comparison

Histograms and cumulative distribution function (CDF) plots were generated for the reduced subset of realizations obtained using the proposed method and the complete superset of realizations. The reduced subset of realizations obtained using the proposed method was compared to the reduced subset of realizations obtained from the traditional ranking method, kernel k-means clustering method and random selection method.

Figures 4 and 5 present the histogram and CDF plot showing the NPV distribution of the original superset of 100 realizations and the selected subset of 10 realizations using the proposed method, respectively. The shapes of the histogram in Fig. 4 between the superset of realizations and the selected subset of realizations obtained using the proposed method confirms similarities in the statistical characteristics between the two distributions. In comparison, the histograms of the selected realizations obtained by k-means clustering, K_{net} -based ranking and random selection have a significantly different shape than the histogram from the superset. In addition to the CDF plot of the proposed method, Fig. 5 also has the CDF plots of the selected realizations obtained by random selection, k-means clustering and K_{net} -based ranking. It is evident from the CDF plots that the selected subset of realizations using the proposed method contains the realizations which have the maximum and minimum NPV values among the superset of realizations. Distributions obtained from the subset of realizations using random selection and k-means clustering are a poor representation of the superset.

Figure 6 shows the COP versus time plot for 10 realizations selected from the proposed method, kernel k-means clustering, random selection and K_{net} ranking.

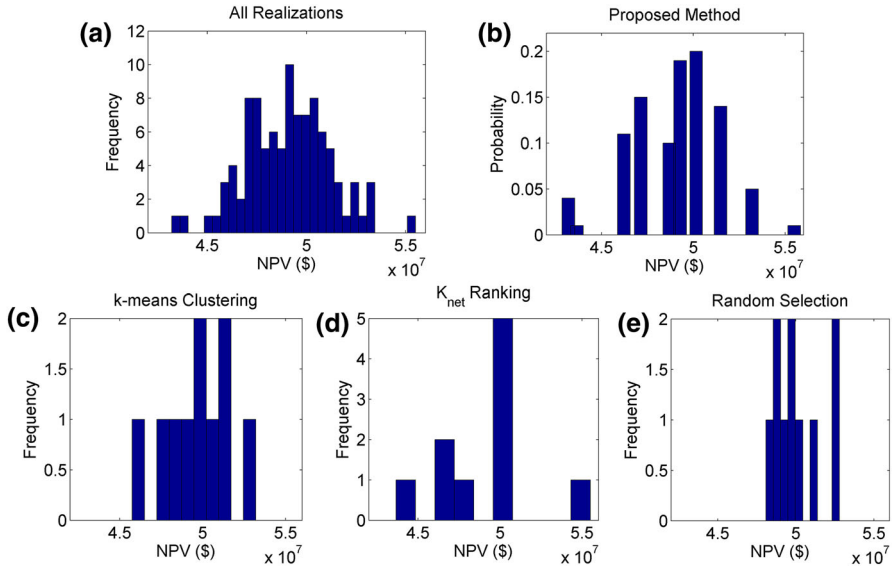


Fig. 4 Histograms using NPV for **a** a superset of 100 realizations (*top*) and 10 selected realizations using **b** the proposed method, **c** kernel k-means clustering, **d** K_{net} ranking and **e** random selection

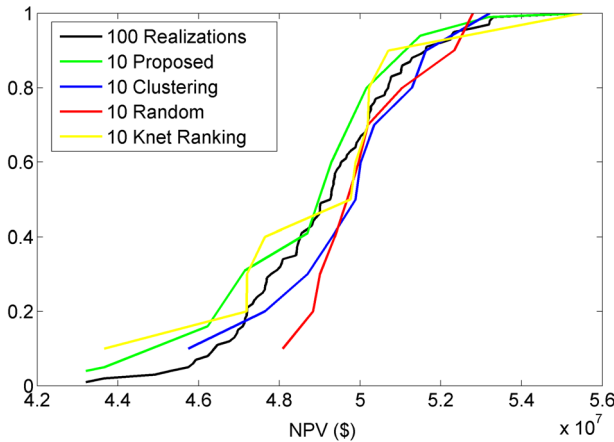


Fig. 5 CDF plot comparison

Plots for the reduced realizations are all superimposed over the COP versus time plots for all 100 realizations from the original set of realizations. Figure 6 confirms that the selected subset of realizations from the proposed method generates a distribution that is a good representation of the distribution of the original superset as the selected realizations evenly cover the entire range of the original superset of realizations.

To check the quality of the selected realizations, expected production parameters are also calculated for comparison. The expected COP is the sum of all of the products

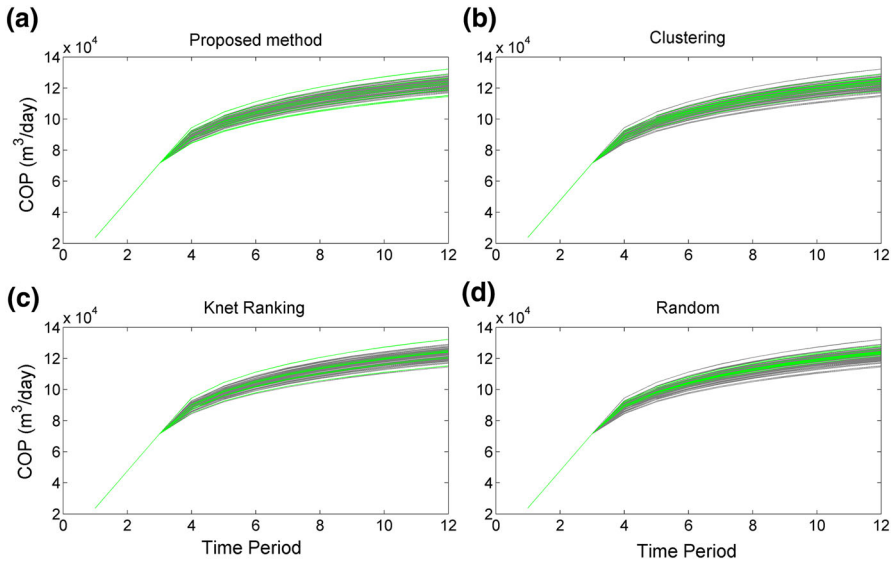


Fig. 6 COP versus time plot for 10 realizations selected from **a** the proposed method, **b** kernel k-means clustering, **c** K_{net} ranking and **d** random selection

between the probability of selecting a realization and the corresponding COP values of that realization as given by the following equation

$$COP_{exp} = \sum_i COP_i \cdot p_i, \tag{27}$$

where p_i is the probability of a selected realization i and COP_i is the corresponding COP value of that selected realization. The expected COP for the subset of realizations obtained using the proposed method is calculated using Eq. (27). All of the selected realizations using the random selection, clustering and static measure-based ranking methods have equal probabilities of being selected, and therefore, the expected COP for the subset of realizations is the average of the COP values for the selected realizations. Using a similar idea, the expected NPV is calculated using Eq. (28)

$$NPV_{Expected} = \sum_i NPV_i \cdot p_i, \tag{28}$$

where p_i is the probability of a selected realization i and NPV_i is the corresponding NPV value. The expected COP and NPV were calculated for the selected subset of realizations using different methods. The maximum, minimum and expected NPV and COP values for the different realization reduction methods are given in Table 5. The proposed method selects realizations to preserve the characteristics of the superset, and this is verified by the small difference between the expected NPV value of the selected realizations by the proposed method (i.e., 4.901×10^7) and the expected NPV from all of the realizations (i.e., 4.907×10^7). The expected COP from the selected subset of realizations using the proposed method (1.229×10^7) is also very close to the

Table 5 Reservoir simulation results of all realizations and selected realizations

	NPV _{max} ($\times 10^7$)	NPV _{min} ($\times 10^7$)	NPV _{exp} ($\times 10^7$)	COP _{max} ($\times 10^5$)	COP _{min} ($\times 10^5$)	COP _{exp} ($\times 10^5$)
All realizations	5.550	4.321	4.907	1.322	1.145	1.230
Proposed method	5.550	4.321	4.901	1.322	1.145	1.229
Ranking						
K_{net}	5.550	4.366	4.920	1.322	1.151	1.231
ϕ_{net}	5.550	4.366	4.941	1.322	1.151	1.234
$S_{iw,net}$	5.550	4.366	4.958	1.322	1.151	1.237
F_{net}	5.550	4.366	4.900	1.322	1.151	1.229
F_{LC}	5.550	4.491	5.007	1.322	1.169	1.244
PV_{net}	5.550	4.366	4.895	1.322	1.151	1.228
OOIP	5.550	4.366	4.926	1.322	1.151	1.232
$O O I P_{net}$	5.550	4.366	4.895	1.322	1.151	1.228
CHV _{local}	5.550	4.491	5.007	1.322	1.169	1.244
Clustering	5.321	4.575	4.978	1.289	1.182	1.240
Random	5.279	4.809	5.013	1.283	1.215	1.245

NPV_{max}, NPV_{min} and NPV_{exp} are the maximum, minimum and expected NPV from all/selected realizations, respectively; COP_{max}, COP_{min} and COP_{exp} are the maximum, minimum and expected COP from all/selected realizations, respectively

expected COP from the superset of realizations (1.230×10^7). The proposed method leads to results containing the realization with the maximum and minimum NPV and COP values from the original full set of realizations as can be seen in Table 5. The other realization reduction methods' results generally do not contain the realizations representing the maximum and minimum NPV and COP values.

The plots showing the expected COP values for the selected subset of realizations at different time periods using the proposed method, clustering method, random selection and ranking method are shown in Fig. 7. A magnified part of the plot around the tenth time period shows that the expected COP of the selected realizations using the proposed method is closest to the expected COP of the full set of realizations. The expected COP values of the subset of realizations selected using random selection and kernel k-means clustering significantly deviate from the expected COP of the full set of realizations. Compared to the single static measure-based ranking method, the proposed method provides an expected COP value which is closer to that of the original superset of realizations.

The water cut versus the time period plot for the selected realizations using different realization reduction methods is given in Fig. 8. To see how similar the distribution of the reduced realization is to that of the original superset of realizations, the water cut plots of selected realizations are superimposed over the water cut plots of all of the realizations in the superset. A magnified version of the water cut plots around the fourth time period is provided in Fig. 9 to show the details. The water cut plots confirm that the reduced subset of realizations from the proposed method is a good representation of the superset of realizations. Specifically, the maximum and minimum water cuts at

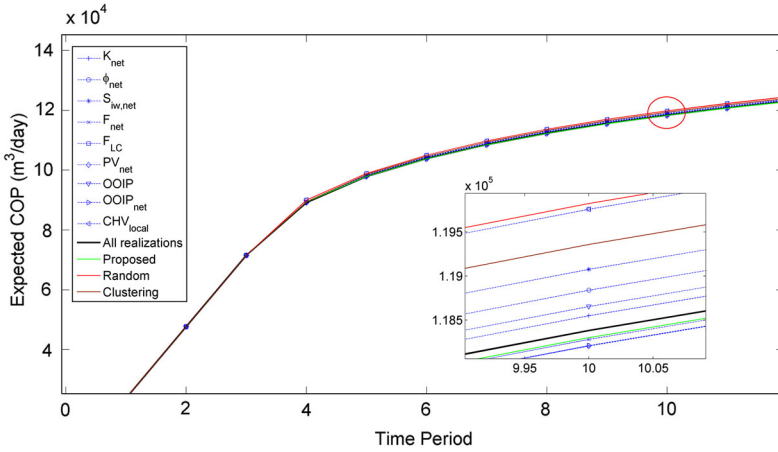


Fig. 7 Expected COP plots of the different realization reduction methods; the figure inside is a magnified version of the plot to show the details

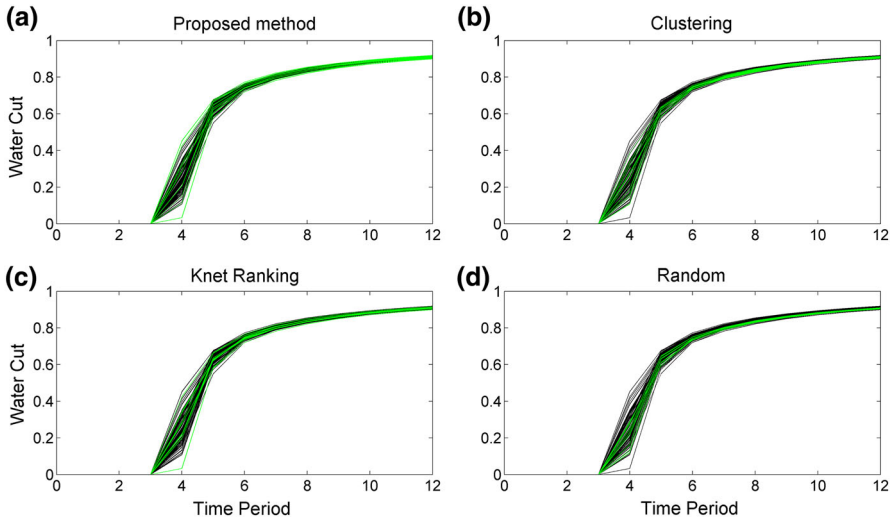


Fig. 8 Water cut plots for 10 realizations selected from **a** the proposed method, **b** kernel k-means clustering, **c** K_{net} ranking and **d** random selection. The *bottom figure* is a magnified version of the plots to show the water breakthrough time

each time period are recovered very well by the proposed method, while the selected realizations using random selection, k-means clustering or K_{net} -based ranking do not cover the entire range of water cut plots for all of the realizations in the superset. It is evident from the magnified part of the water cut plot that the proposed method incorporates realizations which have the highest and lowest water breakthrough time among the superset of realizations.

The results demonstrate that the proposed method generates a subset of realizations which gives a good representation of the superset of realizations in local criterion such as water cut.

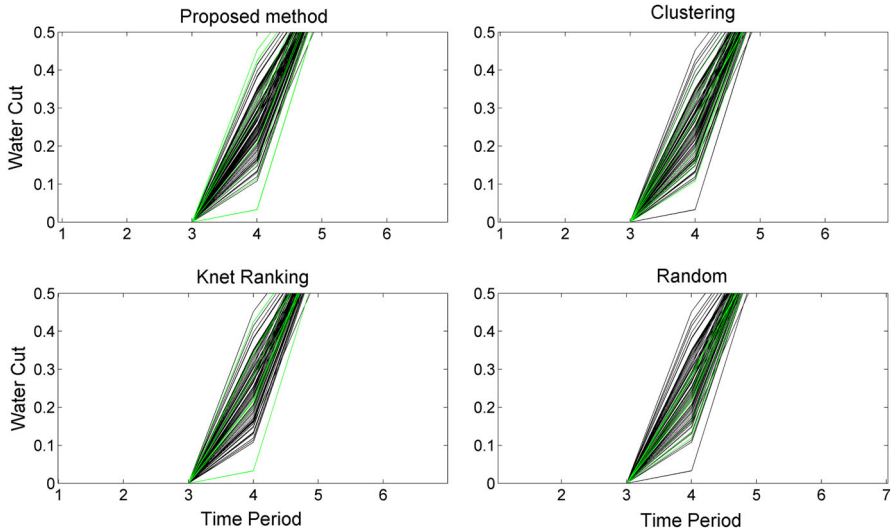


Fig. 9 Water cut profiles around the fourth period

5 Conclusions

In this study, a mixed-integer linear optimization model is proposed to reduce the geological uncertainty in reservoir simulations by selecting a small subset of realizations from a larger superset of realizations. The proposed realization reduction method minimizes the probability distance between the original distribution of the superset of realizations and that of the reduced subset of realizations. The proposed method not only selects a smaller subset of realizations, but also assigns a new probability to each of those realizations. The proposed method is efficient and computationally inexpensive compared to other simplified flow-based methods. As a result, the proposed method is a good candidate for complex realistic reservoirs. The case study demonstrated that the selected realizations from the proposed method have very close statistical characteristics to the distribution of the superset. The proposed method leads to a distribution which is a better representation of the original distribution of all of the realizations than the distributions from the ranking method, kernel k-means clustering method or random selection. The selected realizations from the proposed method have a good coverage of the superset of realizations in terms of the maximum, minimum and expected performance. The proposed method will be very useful for quantifying uncertainty in reservoir performance and reservoir development decision-making.

Acknowledgments The authors would like to acknowledge the financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant Program.

Appendix A: Result of the traditional ranking method using single static measure

Table 6 Results for selected realizations using static measure-based traditional ranking method

Rank K_{net}	ID	80	24	43	83	65	19	50	51	93	64
Measure		0.110095	0.113903	0.114924	0.11577	0.116863	0.118037	0.119597	0.120212	0.122409	0.131013
NPV		43,664,800	47,640,100	49,763,600	47,196,900	47,191,500	49,887,900	50,219,700	50,191,100	50,699,500	55,497,500
COP		115,143	120,888	123,958	120,248	120,240	124,137	124,617	124,575	125,310	132,245
Rank ϕ_{net}	ID	80	24	43	85	65	19	50	51	93	64
Measure		0.311739	0.314379	0.315108	0.315591	0.316291	0.317081	0.318101	0.318519	0.319887	0.325338
NPV		43,664,800	47,640,100	49,763,600	49,291,000	47,191,500	49,887,900	50,219,700	50,191,100	50,699,500	55,497,500
COP		115,143	120,888	123,958	123,274	120,240	124,137	124,617	124,575	125,310	132,245
Rank $S_{1w,net}$	ID	80	24	43	72	96	19	87	51	93	64
Measure		0.69326	0.694931	0.695405	0.695669	0.696057	0.696555	0.697178	0.697428	0.69821	0.701384
NPV		43,664,800	47,640,100	49,763,600	49,652,100	47,135,900	49,887,900	51,628,000	50,191,100	50,699,500	55,497,500
COP		115,143	120,888	123,958	123,796	120,160	124,137	126,652	124,575	125,310	132,245
Rank F_{net}	ID	80	68	56	62	45	54	77	31	39	64
Measure		0.035909	0.222848	0.268985	0.335848	0.392773	0.440894	0.502288	0.543212	0.608697	0.878879
NPV		43,664,800	45,863,600	47,688,000	48,427,800	48,545,900	48,458,600	50,192,000	50,381,300	51,238,200	55,497,500
COP		115,143	118,321	120,958	122,027	122,195	122,071	124,577	124,850	126,089	132,245
Rank F_{LC}	ID	1	17	33	44	61	74	90	5	57	64
Measure		0	0	0	0	0	0	0	0.307818	0.535045	0.878833
NPV		52,264,700	48,021,100	50,265,300	49,679,300	523,27,100	50,131,200	44,909,400	47,378,400	50,267,000	55,497,500
COP		127,572	121,439	124,683	123,836	127,663	124,489	116,941	120,510	124,685	132,245

Table 6 continued

Rank PV_{net}	ID	80	68	56	94	67	54	77	31	39	64
Measure		738,822	4641.51	5580.93	7031.45	8198.67	9263.33	10601.7	11416.1	12870.2	18871.5
NPV		43,664,800	45,863,600	47,688,000	48,411,600	48,086,500	48,458,600	50,192,000	50,381,300	51,238,200	55,497,500
COP		115,143	118,321	120,958	12,2003	121,534	12,2071	124,577	124,850	126,089	132,245
Rank OOI _P	ID	80	69	38	15	67	54	77	30	98	64
Measure		6095.97	6185.46	6200.06	6218.59	6235.06	6248.6	6266.79	6275.41	6291.19	6382.59
NPV		43,664,800	47,095,300	46,945,900	49,003,700	48,086,500	48,458,600	50,192,000	50,347,000	53,302,800	55,497,500
COP		115,143	120,101	119,885	122,859	121,534	122,071	124,577	124,801	129,073	132,245
Rank $OOI_{P_{net}}$	ID	80	68	56	62	67	54	77	31	39	64
Measure		226,545	1411.89	1701.82	2133.68	2491.19	2802.07	3197.65	3452.49	3877.44	5630.32
NPV		43,664,800	45,863,600	47,688,000	48,427,800	48,086,500	48,458,600	50,192,000	50,381,300	51,238,200	55,497,500
COP		115,143	118,321	120,958	122,027	121,534	122,071	124,577	124,850	126,089	132,245
Rank CHV_{local}	ID	1	17	33	44	61	74	90	5	57	64
Measure		0	0	0	0	0	0	0	1948.46	3399.02	5630.03
NPV		52,264,700	48,021,100	50,265,300	49,679,300	52,327,100	50,131,200	44,909,400	47,378,400	50,267,000	55,497,500
COP		127,572	121,439	124,683	123,836	127,663	124,489	116,941	120,510	124,685	132,245

Appendix B. Proposed optimization model for realization reduction

The complete optimization model for realization reduction is given as follows

$$\min D_{Kan} = \sum_{i \in I} p_i^{orig} d_i.$$

Subject to

$$\begin{aligned} d_i &= \sum_{i' \in I} c_{i,i'} v_{i,i'} \forall i \in I \\ \sum_{i' \in I} v_{i,i'} &= y_i \forall i \in I \\ 0 \leq v_{i,i'} &\leq 1 - y_{i'} \forall i, i' \in I \\ \sum_{i \in I} y_i &= R \\ \sum_{i \in I_{SB}} (1 - y_i) &\geq 2 \\ \sum_{i \in I_{SW}} (1 - y_i) &\geq 2 \\ p_{i'}^{new} &= (1 - y_{i'}) p_{i'}^{orig} + \sum_i v_{i,i'} p_i^{orig} \forall i' \in I \\ y_i &\in \{0, 1\} \forall i \in I. \end{aligned}$$

Input Parameters

- R the total number of realizations to be removed
- p_i^{orig} the original probabilities of realizations, i , is normally set as equal to $p_i^{orig} = 1/|I|$, where $|I|$ is the size of the set I
- $c_{i,i'}$ the distance between two geological realizations

Variables

- y_i binary variables which denote whether a realization is removed ($y_i = 1$) or not ($y_i = 0$)
- $v_{i,i'}$ continuous variables ($v_{i,i'} \in [0, 1]$) which denote the fraction of its probability mass that is transported from realization i to realization i'
- p_i^{new} continuous variables ($p_i^{new} \in [0, 1]$) which denote the new probabilities of realizations i after optimal probability mass transportation. Notice that if $p_i^{new} = 0$, it means that realization i is removed

References

Armstrong M, Ndiaye A, Razanatsimba R, Galli A (2013) Scenario reduction applied to geostatistical simulations. *Math Geosci* 45:165–182

- Ballin PR, Journel AG, Aziz K (1992) Prediction of uncertainty in reservoir performance forecast. *JCPT* 31(4):52–62
- Christie MA, Blunt MJ (2001) Tenth SPE comparative solution project: a comparison of upscaling techniques. *SPE J* 4(4):308–317
- Deutsch CV (1998) Fortran programs for calculating connectivity of three-dimensional numerical models and for ranking multiple realizations. *Comput Geosci* 24:69–76
- Deutsch CV (1999) Reservoir modeling with publicly available software. *Comput Geosci* 25:355–363
- Deutsch CV, Begg SH (2001) The use of ranking to reduce the required number of realizations. Centre for Computational Geostatistics (CCG) Annual Report 3, Edmonton
- Deutsch CV, Srinivasan S (1996) Improved reservoir management through ranking stochastic reservoir models. In: *SPE/DOE Improved Oil Recovery Symposium*, 21–24 April 1996, Tulsa, USA. doi:[10.2118/35411-MS](https://doi.org/10.2118/35411-MS)
- Dupacova J, Growe-Kuska N, Romisch W (2003) Scenario reduction in stochastic programming an approach using probability metrics. *Math Progr* 95:493–511
- Fenik DR, Nouri A, Deutsch CV (2009) Criteria for ranking realizations in the investigation of SAGD reservoir performance. In: *Proceedings of Canadian International Petroleum Conference*, 16–18 June 2009, Calgary, Canada. doi:[10.2118/2009-191](https://doi.org/10.2118/2009-191)
- Gilman JR, Meng H-Z, Uland MJ, Dzurman PJ, Cosic S (2002) Statistical ranking of stochastic geomodels using streamline simulation: A field application. In: *SPE Annual Technical Conference and Exhibition*, 29 Sept–2 Oct 2002, San Antonio, USA. doi:[10.2118/77374-MS](https://doi.org/10.2118/77374-MS)
- Kantorovich LV (1942) On the transfer of masses. *Dokl Akad Nauk* 37:227–229
- Li SH, Deutsch CV, Si JH (2012) Ranking geostatistical reservoir models with modified connected hydrocarbon volume. In: *Ninth International Geostatistics Congress*, 11–15 June 2012, Oslo, Norway
- Li Z, Floudas CA (2014) Optimal scenario reduction framework based on distance of uncertainty distribution and output performance: I. Single reduction via mixed integer linear optimization. *Comput Chem Eng* 70:50–66
- Lie KA, Krogstad S, Ligaarden I, Natvig J, Nilsen H, Skaflestad B (2012) Open-source MATLAB implementation of consistent discretisations on complex grids. *Comput Geosci* 16:297–322
- IBM (2009) User's Manual for CPLEX V12.1. [ftp://public.dhe.ibm.com/](http://public.dhe.ibm.com/)
- McLennan JA, Deutsch CV (2005) Selecting geostatistical realizations by measures of discounted connected bitumen. In: *SPE International Thermal Operations and Heavy Oil Symposium*, 1–3 Nov 2005, Calgary, Canada. doi:[10.2118/98168-MS](https://doi.org/10.2118/98168-MS)
- McLennan JA, Deutsch CV (2004) SAGD reservoir characterization using geostatistics: applications to the Athabasca Oil Sands, Alberta, Canada. In: *Press, Canadian Heavy Oil Association Handbook*
- Monge G (1781) *Mémoire sur la théorie des déblais et des remblais: Histoire de l'Académie Royale des Sciences*
- Park K, Caers J (2011) *Metrel: Petrel plug-in for modelling uncertainty in metric space*. Stanford University, Department of Energy Resource Engineering
- Scheidt C, Caers J (2009) Representing spatial uncertainty using distances and kernels. *Math Geosci* 41:397–419
- Scheidt C, Caers J (2010) Bootstrap confidence intervals for reservoir model selection techniques. *Comput Geosci* 14:369–382
- Singh AP, Maucec M, Carvajal GA, Mirzadeh S, Knabe SP, Al-Jasmi AK, El Din IH, (2014) Uncertainty quantification of forecasted oil recovery using dynamic model ranking with application to a ME carbonate reservoir. In: *International Petroleum Technology Conference*, 20–22 Jan 2014, Doha, Qatar. doi:[10.2523/17476-MS](https://doi.org/10.2523/17476-MS)
- Shi J, Malik J (2000) Normalized-cut and image segmentation. *IEEE Trans Pattern Anal Mach Intel* 22(8):888–905