RMFinder 2.0: An Improved Interactive Multi-criteria Scenario Reduction Methodology

Luis A. A. Meira, Guilherme P. Coelho, Celmar G. Silva, School of Technology (FT), University of Campinas (Unicamp); Antonio Alberto S. Santos, Denis J. Schiozer, Center for Petroleum Studies (CEPETRO), University of Campinas (Unicamp)

Abstract

This paper presents an extension of the RMFinder technique, previously proposed to identify representative models (RMs) within the decision-making process in oil fields. As there are several uncertainties associated with this decision-making process, a large number of scenarios are supposed to be analyzed, so that high-quality production strategies can be defined. Such broad analysis is often unfeasible, so techniques to automatically identify RMs are particularly relevant. The original RMFinder does not consider individual probabilities of each RM, which may not be accurate when the risk curves of the problem are estimated. Therefore, a mechanism to calculate the individual probability of each RM was developed here, together with a graphical way to visualize different proposals of RMs. To automatically identify the optimal probability of each RM, this new version of RMFinder minimizes the deviation between the risk curves generated with the selected RMs and the original risk curves of the problem. The graphical approach automatically exhibits, in a single page per solution, the RM dispersion in the scatter plots, the resulting risk curves and the differences between attribute-level distributions. This helps the decision makers to visualize and compare different sets of RMs. The proposed methodology was applied to a small synthetic problem and to three reservoir models based on real-world Brazilian fields: (i) UNISIM-I-D, a benchmark case based on the Namorado field; (ii) UNISIM-II-D, a benchmark case based on a highly fractured pre-salt carbonate reservoir; and (iii) ST001a, a highly heterogeneous heavy oil offshore field. The obtained sets of RMs were evaluated by experts and considered appropriate to the studied problems, being adopted as the standard models in the following steps of the decision-making process to define the production strategies under uncertainties.

Introduction

One of the main goals of the decision making process in oil fields is a proper definition of the production strategy\(^1\) to be implemented. The basis of the production strategy definition procedure is the geological model of the field, which is used by computational simulators to estimate the behavior of the reservoir during the concession time (given a production strategy). Therefore, given the behavior of the reservoir, it is possible to evaluate the quality of the production strategy. Notice that the production strategies can be evaluated according to different criteria, such as the Net Present Value (NPV), for example. In this context, geological models of oil fields usually contain

\(^1\) The production strategy is composed by the infrastructure of wells, pipes, platforms and manifolds, for example.
uncertainties that must be properly considered by the decision maker, since the simulation results of a particular production strategy may be different when two distinct values of an uncertain variable of the model are considered.

Ideally, the decision maker should consider all possible combination of values of the uncertain attributes when defining a production strategy, which may lead to thousands of different scenarios to be analyzed. The simulation of the behavior of the oil field for each possible production strategy and for each scenario is unfeasible in real-world situations, as each simulation may often require several hours, even days, to conclude. Therefore, scenario reduction techniques are mandatory, so that decision makers can work with a feasible, representative set of models of the whole problem.

In this context, the RMFinder technique (Meira et al., 2016) is an optimization-based approach proposed to automatically identify, according to the user’s preferences, subsets of representative models (RMs) of oil fields. In Meira et al. (2016), the mathematical function to be optimized was developed considering the representativeness of a given subset of models with respect to the cross-plots of the main output variables, the associated risk curves and the probability distribution of the uncertain attribute with levels of the problem.

However, although RMFinder has been presenting superior results when compared to previous manual approaches to identify RMs (Meira et al., 2016) and has been used in real-world applications (Morosov and Schiozer, 2016), it does not indicate to the user the individual probability of each representative model. This may be an issue in situations in which the new risk curves of the problem, based on the RMs, are required. In this paper, we have verified that the naive approach of considering all the RMs equiprobable may lead to distortions on the risk curves. Therefore, an optimization-based approach was developed and integrated into RMFinder to estimate the correct individual probabilities of each RM.

Another improvement was the inclusion of a new way to visualize and compare RMFinder’s results (sets of RMs). Such new organization of the results was developed according to Information Visualization techniques (Mazza, 2009), and aimed at allowing the user to easily verify the quality of the set of RMs returned by RMFinder and even compare the results obtained from two different runs of the algorithm. The new version of RMFinder is referred in this paper as RMFinder 2.0.

RMFinder 2.0 was applied to a small synthetic problem and to three problems based on real-world oil fields: UNISIM-I-D (Gaspar et al., 2015), which is benchmark case based on the Namorado field (Brazil), UNISIM-II-D (Correia et al., 2015), also a benchmark case but based on a highly fractured pre-salt carbonate reservoir, and the problem that will be called here ST001a, which is a highly heterogeneous heavy oil offshore field. The obtained results were analyzed by experts, considered adequate and adopted as the standard models in the remaining steps of the decision-making process to define the best production strategies for each problem.

With the improvements included into RMFinder 2.0, which will be presented here, a practicing engineer can not only identify a set of representative models for a given problem, but also visually validate such RMs and identify a feasible number of RMs that leads to the best representation of the original uncertainties of the problem. Besides, given that the selection of RMs can be made within minutes, this task can be repeated and refined whenever new information about the problem becomes available.

This paper is organized as follows. First, the theoretical background required for the proposals of this paper will be discussed, followed by a literature review on related works. Then, the improvements proposed here for RMFinder will be thoroughly discussed. The experimental methodology and results will follow, together with some final conclusions.

**Theoretical Background**

This section presents the main theoretical aspects related to this work. First, the task of selecting representative models is contextualized within the decision making process in oil fields and some basic definitions are provided. Then, the original RMFinder, which is the basis of the improved technique proposed here, is briefly presented. And finally, the main concepts of information visualization adopted in this work are discussed.

**Selection of Representative Models**

As previously mentioned, the strategy production definition of hundreds, even thousands of different scenarios is generally unfeasible in practice. Therefore, decision makers often work with a small subset of scenarios that properly represent the properties of the original problem, so that the simulation times become viable. In this context, there are several approaches in the literature to identify the representative models of a given problem (Ligero et al., 2005; Scheidt and Caers, 2009; Sarma et al., 2013; Armstrong et al., 2013; Meira et al., 2016), each one based on different criteria.

Strategies to select representative models are often used within broader decision-making methodologies so the number of scenarios can be reduced before further analyses. For example, the original RMFinder was developed for the 12-step methodology proposed by Schiozer et al. (2015). The 12 steps of such methodology can be summarized as follows:
1. Characterize the reservoir (build models, develop scenarios and estimate probabilities).
2. Select a base case (scenario combined with a production strategy) to build and calibrate the simulation model. This case is named “Base 0”.
3. Using data collected from wells already in the field, search for inconsistencies in “Base 0”.
4. After the discretization of continuous variables of the problem into levels (see below), use the probability of each level to generate a set of scenarios for the problem.
5. Using production data from the wells already in the field, verify the feasibility of each scenario and eliminate the unfeasible ones. From this reduced set of scenarios, build a new base case (named “P50”).
6. Identify a production strategy for “P50”.
7. With the production strategy identified in Step 6, estimate the risk curves (see below) considering all the scenarios obtained in Step 5.
8. Further reduce the set of scenarios obtained in Step 5, by identifying a set of representative models (RMs).
9. Obtain production strategies for each RM identified in Step 8;
10. Identify the best strategy under uncertainty that combines all possible strategies, using a risk-return analysis.
11. Try to improve the strategy identified in Step 10, improving its robustness and flexibility.
12. Generate the final risk curves and perform the decision analysis.

As mentioned in Step 4 above, each continuous uncertain attribute \(a_i\), associated with the decision-making process is generally discretized into \(n_a\) levels, and each of these attribute levels have a particular probability \(p_{a,i}\) associated with it (being \(l\) the level of attribute \(a\)). Given that a particular scenario is obtained with the definition of specific values for all its attributes, if all uncertain attributes are discrete (or were discretized), the total number of scenarios in a particular problem is given by Eq. (1).

\[
|\tau| = \prod_{a \in A} n_a \tag{1}
\]

where \(\tau\) is the set of scenarios, \(A\) the set of attributes of the problem and \(|\cdot|\) is the cardinality operator. Eq. (1) shows that the number of scenarios grows significantly with both the number of attributes and the number of levels. To illustrate, a small problem with 10 uncertain attributes discretized into 3 attribute-levels would result in \(|\tau| = 3^{10} = 59,049\) different scenarios to be analyzed.

The RMFinder representativeness concept is based on three main aspects. The first aspect that should be considered when selecting a subset \(\tau'\) of representative models from \(\tau\), is to maintain the proportion of attribute levels in \(\tau'\) as close as possible to that observed in \(\tau\).

The second aspect is associated with the production variables of problem, which could be the Net Present Value (NPV), Oil Recovery Factor (ORF), Oil Production (Np) and Water Production (Wp), among others. It is often required that the subset of RMs properly represent the distribution of the original set of models in the domain of the production variables.

For example, if three production variables (say NPV, ORF and Np) are considered for a given production strategy \(Pr\), the original set of scenarios can be plotted in \(\mathbb{R}^3\). Therefore, the set of RMs should be selected so that they adequately represent the original “cloud” of points in \(\mathbb{R}^3\).

Finally, the third aspect that should be considered are the risk curves, which correspond to plots of each production variable \(f(s)\) vs. its cumulative probability \(p_c^f: \mathbb{R} \to [0, 1]\), given in Eq. (2).

\[
p_c^f(s, \tau) = \frac{p(s, \tau)}{2} + \sum_{s' \in f(s, Pr) \cap f(s, \tau)} p(s', \tau) \tag{2}
\]

where \(\tau\) is the set of scenarios being considered, \(Pr\) is a production strategy, \(s\) is a scenario, \(f(s, Pr)\) is the value of the production variable given \(s\) and \(Pr\) (obtained through computational simulation), and \(p(s, \tau)\) is the probability of each scenario \(s\) in \(\tau\). In other words, a cumulative probability \(p_c^f(s, \tau) = y\) indicates that, for the production strategy \(Pr\) and given a set of scenarios \(\tau\), there is a probability \(y\) that the final value of the production variable \(f\) is greater than or equal to \(f(s, Pr)\).

Figure 1 represents a typical risk curve for a particular production variable (NPV). In this curve, scenarios (points) with lower values of NPV (and higher values of \(p_c^{NPV}\)) are known as pessimistic, while scenarios with higher values of NPV (and lower values of \(p_c^{NPV}\)) are optimistic. Therefore, when selecting representative models, such RMs must also properly represent the original risk curves of all production variables of the problem.
The original RMFinder
Considering the 12-step methodology presented in the previous section (Schiozer et al., 2015), RMFinder was originally proposed to automate Step 8. In other words, it is supposed to receive the original (full) set of feasible models of the problem and return a subset of models that properly represent the original set (the RMs). In this process, the user is responsible for defining the desired number of RMs together with a set of weights, which balance the importance of different aspects that compose the “representativeness” concept adopted by RMFinder (Meira et al., 2016).

This concept of representativeness of a set of RMs is defined by a mathematical function comprised of three components (criteria): (i) $F_{cross}(r')$, which is responsible for the representativeness with respect to the production variables; (ii) $F_{risk}(r')$, which is associated with the risk curves of the problem; and (iii) $F_{atr}(r')$, related to the proportion of attribute with levels. These three components are combined into a single mathematical function, $F(r')$, through a weighted sum with weights defined by the user. In this context, the set of RMs with highest representativeness is the one that minimizes the mathematical function $F(r')$, given in Eq. (3).

$$F(r') = q_1 \cdot F_{cross}(r') + q_2 \cdot F_{risk}(r') + q_3 \cdot F_{atr}(r')$$ (3)

To obtain the set of RMs, Meira et al. (2016) suggest the use of meta-heuristics to perform this optimization task and adopted the immune-inspired algorithm named cob-alNeC (Coelho et al., 2011) in their experiments. However, any meta-heuristics can be employed to perform such optimization.

Equation (3) indicates two of the main characteristics of RMFinder: (i) it is a multi-criteria approach to identify representative models, since it combines three components into its main objective function; and (ii) it is an interactive approach, as the user has a fundamental role in the process through the definition of the weights $q_i$, which transmits to the system the user’s preferences.

This section is not intended to provide all the details of the original RMFinder, but only to discuss those key aspects directly related to the improvements proposed here. Therefore, only the components of the main objective function $F(r')$ will be discussed. For further details, the reader is referred to the work of Meira et al. (2016).

The first component of $F(r')$ is the cross-plot objective function ($F_{cross}(r')$), which is intended to evaluate the representativeness of the set of RMs with respect to the production variables of the problem. In the original RMFinder, four production variables were considered (NPV, ORF, Np and Wp) and, for each scenario, the production strategy defined for a base model is simulated and the values of the production variables obtained. These four values of the production variables allow the representation of each scenario as a point in $\mathbb{R}^4$, so, ideally, the set of RMs should properly represent this “cloud” of points in $\mathbb{R}^4$. However, RMFinder was developed to mimic the manual process of identification of RMs so, instead of considering the $\mathbb{R}^4$ space, it works with six projections of the cloud of points in $\mathbb{R}^2$, which are named cross-plots. This projections are $NPV \times ORF$, $NPV \times Np$, $NPV \times Wp$, $ORF \times Np$, $ORF \times Wp$ and $Wp \times Np$. For each cross-plot, the set of RMs should be uniformly spread
over the original cloud of points, representing the distribution of the original scenarios. Therefore, for a set of RMs \( \tau' \), \( \mathcal{F}_{\text{cross}}(\tau') \) is given by Eq. (4).

\[
\mathcal{F}_{\text{cross}}(\tau') = q_1^c \cdot \mathcal{F}_{\text{NPV,DRP}}(\tau') + q_2^c \cdot \mathcal{F}_{\text{NPV,WP}}(\tau') + q_3^c \cdot \mathcal{F}_{\text{ORF,WP}}(\tau') + q_4^c \cdot \mathcal{F}_{\text{ORF,NP}}(\tau') + q_5^c \cdot \mathcal{F}_{\text{cross}}(\tau') + q_6^c \cdot \mathcal{F}_{\text{cross}}(\tau'),
\]

where \( q_i^c \in \mathbb{R}_+ \) are weights that must be set by the user (according to his/her preferences, indicating which cross-plots are more important in the process) and \( \mathcal{F}_{\text{cross}}(\tau') \) is the squared-error distortion for the \( f \times g \) cross-plot. Such squared-error distortion for a given cross-plot is basically the sum of the smallest squared Euclidean distances between each scenario and the closest RM.

The *risk-curve* objective function, \( \mathcal{F}_{\text{risk}}(\tau') \), considers the risk curves for the same previously mentioned production variables, and its minimum value is obtained when the RMs are selected in a way that they become equally spaced on all original risk curves of the problem. Given that there are four risk curves, \( \mathcal{F}_{\text{risk}}(\tau') \) has four components, as presented in Eq. (5).

\[
\mathcal{F}_{\text{risk}}(\tau') = q_1^r \cdot \mathcal{F}_{\text{NPV}}(\tau') + q_2^r \cdot \mathcal{F}_{\text{ORF}}(\tau') + q_3^r \cdot \mathcal{F}_{\text{NPV}}(\tau') + q_4^r \cdot \mathcal{F}_{\text{ORF}}(\tau'),
\]

where \( q_i^r \in \mathbb{R}_+ \) are weights that must be set by the user and can be used to increase or decrease the importance of a good distribution of the RMs in a particular risk curve. Each \( \mathcal{F}_{\text{risk}}(\tau') \) evaluates the spread of the RMs in risk curve \( f \). To calculate \( \mathcal{F}_{\text{risk}}(\tau') \), the RMs in \( \tau' \) are sorted with respect to their values of production variable \( f \) and the squared Euclidean distance between each RM is calculated and summed. Meira et al. (2016) demonstrated that \( \mathcal{F}_{\text{risk}}(\tau') \) will be minimum if the RMs in \( \tau' \) are equally spread on the risk curve of production variable \( f \).

Finally, the last component of the mathematical function \( \mathcal{F}(\tau') \) is the *attribute-level* objective function \( \mathcal{F}_{\text{atr}}(\tau') \), which evaluates how different the distribution of attribute levels is on the set of RMs \( \tau' \) when compared to the original set of scenarios. \( \mathcal{F}_{\text{atr}}(\tau') \) is given in Eq. (6).

\[
\mathcal{F}_{\text{atr}}(\tau') = (1 + \mathcal{F}_{\text{penalty}}(\tau')) \cdot \mathcal{F}_{\text{dist}}(\tau'),
\]

where \( \mathcal{F}_{\text{penalty}}(\tau') \) is a binary penalty function and \( \mathcal{F}_{\text{dist}}(\tau') \) corresponds to the squared sum of the differences observed between the relative frequencies of each attribute level in \( \tau' \) and in the original set of scenarios. \( \mathcal{F}_{\text{dist}}(\tau') \) evaluates the difference between the distribution of attribute levels in \( \tau' \) and in the original set of scenarios, while \( \mathcal{F}_{\text{penalty}}(\tau') \) penalizes sets of RMs that do not contain at least one occurrence of all attribute levels (it returns 0 if all levels of all attributes occur in at least one scenario in \( \tau' \) and 1 otherwise).

The original RMFinder was able to automate the process of RM selection within the 12-step methodology of Schiozer et al. (2015), so the process that used to take days to be solved manually can now be solved within minutes. This created several new possibilities of study, such as the iterative approach adopted by Morosov and Schiozer (2016), who repeated the cycle of (i) data acquisition from the wells, (ii) history matching and (iii) optimization of the production strategy over time. In their work, Morosov and Schiozer (2016) observed that, although better results with respect to the expected monetary value were obtained, the NPV actually decreased after this iterative process was applied to a benchmark problem (UNISIM-I). Such results indicate that companies worldwide must guarantee somehow that their reservoir studies be able to properly predict the reservoir performance.

The three components of RMFinder’s objective function cover the main aspects of representativeness that are expected on a set of representative models, and that is the main advantage of RMFinder. However, such technique does not estimate the individual probabilities of each RM, and this is an important issue, since these probabilities directly influence the shape of the new risk curves. Therefore, in this work a new approach to calculate such probabilities is proposed, together with a new risk-curve component for the mathematical function given in Eq. (3) that evaluates the distance between the risk curves generated with the RMs and the original risk curves of the problem. Besides, the attribute-level component was also divided into two separate components: one responsible for evaluating the differences between the distribution of attribute levels (\( \mathcal{F}_{\text{dist}}(\tau') \)) and the other for ensuring at least one occurrence of each attribute-level in the set of RMs. These new approaches will be discussed in the following sections.

**Information Visualization**

The context of decision-making process in oil fields usually involves large and multidimensional data sets. RMFinder uses graphical representations in order to help system users to analyze these sets. Indeed, this is coherent with the goal of Information Visualization (InfoVis): providing graphical (and possibly interactive) representations of a data set such that they enhance human thinking and reasoning about this set (Card et al.,
have possibly different probabilities, which are not consistent with the original one. New risk curves of the problem.

The original RMFinder, proposed by Meira et al. (2016), did not consider the individual probabilities of each representative model, what corresponds to an indirect assumption that the RMs are equiprobable. Under such conditions, it is possible to apply Eq. (2) to obtain the cumulative probabilities of the RMs and, consequently, the new risk curves of the problem. However, such equiprobability assumption may lead to new risk curves that are not consistent with the original ones. Therefore, in this work we relaxed such assumption so that the RMs can have possibly different probabilities, which allowed us to obtain risk curves more consistent with the original ones.

Related Work

Given the importance of selecting representative models in decision-making processes in oil fields, several strategies have been proposed in the last years. Steagall and Schiozer (2001) focused on the NPV risk curve and proposed that the set of RMs should be constituted by: two models close to P50 (the model with cumulative probability close to 90%), three models close to P50 (the model with cumulative probability close to 50%) and two models close to P10 (the model with cumulative probability close to 10%). To select these models close to P50, P50 and P10, other production variables can be considered by the decision-maker. Schiozer et al. (2004) proposed an extension of the methodology proposed by Steagall and Schiozer (2001), with the additional requirement that the RMs should be representative not only with respect to NPV, but also to Np, Wp and ORF. Sarma et al. (2013) proposed the use of a greedy algorithm to identify the closest RMs to P50, P50 and P10, through the maximization of the attribute-level coverage. Finally, Rahim et al. (2015) proposed a mix-integer linear optimization-based approach to obtain RMs that minimizes the probability distance between the distributions of the original (full) set of scenarios and the RMs.

These previously mentioned strategies can be seen as ranking procedures that try to replicate, in the set of RMs, the representation of the uncertainties observed in the original (full) set of models. However, many of them are based on subjective criteria of the decision makers and demand a time-consuming manual analysis.

Besides the ranking-based approaches, several authors proposed clustering-based techniques, with special contributions mostly to the distance metrics adopted to evaluate the dissimilarity between scenarios. In this context, Suzuki et al. (2008) and Suzuki and Caers (2008) propose the use of the Hausdorff distance, while Caers and Park (2008) and Park (2011) recommend the use of the EnKF metric. Scheidt and Caers (2009) and Armstrong et al. (2013) proposed specific distance metrics, being the former work devoted to apply kernel-based clustering strategies to identify RMs in a synthetic oil reservoir and the latter to apply random search to maximize the NPV in the context of gold mines in Australia. More recently, Shirangi and Durlofsky (2016) introduced a general clustering-based procedure that incorporates flow and permeability based features and defines a low-dimensional flow-response vector to characterize the simulation results.

Given this general context, the original RMFinder can be seen as a ranking procedure, as it indirectly ranks sets of representative models according to its objective function (Eq. (3)), that contains mechanisms from clustering-based approaches, such as the squared-error distortion used in the cross-plots that comes from the well-known k-medoids clustering algorithm.

Proposed Methodology: RMFinder 2.0

The original RMFinder, proposed by Meira et al. (2016), did not consider the individual probabilities of each representative model, what corresponds to an indirect assumption that the RMs are equiprobable. Under such conditions, it is possible to apply Eq. (2) to obtain the cumulative probabilities of the RMs and, consequently, the new risk curves of the problem. However, such equiprobability assumption may lead to new risk curves that are not consistent with the original ones. Therefore, in this work we relaxed such assumption so that the RMs can have possibly different probabilities, which allowed us to obtain risk curves more consistent with the original ones.
To illustrate, let’s consider the UNISIM-I-D case studied by Meira et al. (2016), where RMFinder was used to obtain 9 RMs from the original set of 214 models. Although the discussion here will be based on this particular case, the observations apply to other problems as well.

Meira et al. (2016) simply ignored the individual probability of each RM, assuming that they would be in the same place in the original risk curves of the problem, as illustrated in Figure 2a (only the NPV risk curve will be shown here). However, if the equiprobability assumption is considered (all RMs with probability 1.0/9, given that there are 9 RMs) and the new risk curve is obtained through Eq. (2), we end up with the situation illustrated in Figure 2b: the new risk curve (black dots) is significantly different from the one obtained with the original set of 214 models (purple dots). In this example, the new risk curve clearly presents a pessimistic bias for the models with NPV smaller than that of $P_{50}$ and an optimistic bias for those models with NPV higher than that of $P_{50}$.

When such equiprobability assumption is relaxed and new (possibly different) probabilities are identified for each RM, it is possible to obtain new risk curves in which the RMs are closer to their position in the original risk curves. In other words, the new risk curves, formed by the RMs, tend to become closer to the risk curves obtained with the original models of the problem. This is illustrated in Figure 2c.

![Figure 2](image)

**Figure 2** – NPV risk curves obtained by Meira et al. (2016) for UNISIM-I-D problem. In (a) the probabilities of the RMs (black dots) were ignored, so these models were supposed to be in the same position over the original risk curve (purple dots). In (b), the correct risk curve, calculated with Eq. (2) and assuming equiprobability of the RMs, is presented. And in (c), the new risk curve, considering different probabilities for each RM, is given.

The first attempt to obtain the correct probabilities of the RMs (Attempt 1) was to minimize the sum of the squared differences between the cumulative probability of each RM in the set of RMs and its cumulative probability in the original set of models. To illustrate, the goal of Attempt 1 was to minimize the “black bars” given in Figure 3, but such attempt was not successful.

![Figure 3](image)

**Figure 3** – Illustration of the differences to be minimized (black bars) during the first attempt to obtain the RM probabilities.

When Attempt 1 was implemented, anomalies were observed: some RMs ended up with probability 0.0. To illustrate this aspect consider the three RMs (black dots) obtained for the UNISIM-I-D problem given in Figure 4. When Attempt 1 was run, these three RMs ended up with probabilities 0.5, 0.0 and 0.5 (Figure 4b), which led to cumulative probabilities equal to 0.25, 0.5 and 0.75, respectively, and a perfect adjustment of the new risk curve over the original one, as presented in Figure 4a. However, it is not desirable to indicate a specific model as a pessimist or an optimist, which led to the next approach.

In practice, solutions as those represented in Figure 5, in which the RMs have probabilities different from 0.0, are more suitable. In this case, the probabilities are 0.333, 0.333 and 0.333, and the cumulative probabilities 0.165, 0.5 and 0.835, respectively. Therefore, a new approach (Attempt 2) had to be developed, which considers not only the cumulative probabilities of the RMs, but also their individual probabilities. This new approach...
evaluates the overlap between the risk curves generated by the RMs and by the original models of the problem, and adjusts the individual probabilities of each RM so that the differences between these curves are minimized. Attempt 2 was implemented in RMFinder 2.0.

\begin{figure}[h]
\centering
\begin{subfigure}{0.4\textwidth}
  \includegraphics[width=\textwidth]{figure4a.png}
  \caption{}
\end{subfigure}
\begin{subfigure}{0.4\textwidth}
  \includegraphics[width=\textwidth]{figure4b.png}
  \caption{}
\end{subfigure}
\caption{(a) Illustration of an optimal solution obtained with Attempt 1. In (b), the black bars correspond to the individual probabilities of each RM. Notice that probability 0.0 was assigned to one of the RMs.}
\end{figure}

\begin{figure}[h]
\centering
\begin{subfigure}{0.4\textwidth}
  \includegraphics[width=\textwidth]{figure5a.png}
  \caption{}
\end{subfigure}
\begin{subfigure}{0.4\textwidth}
  \includegraphics[width=\textwidth]{figure5b.png}
  \caption{}
\end{subfigure}
\caption{(a) Illustration of a solution that is worse than the one given in Figure 4, according to Attempt 1’s criteria, but more suitable in practice. In (b), the black bars correspond to the individual probabilities of each RM.}
\end{figure}

Consider a set of scenarios \( T = \{s_1, \ldots, s_n\} \), so that each scenario \( s \in T \) has probability \( p(s, T) = 1/n \) and cumulative probability \( p_c(s, T) \) given by Eq. (2). Also consider a function \( f^* = (f_1, \ldots, f_k) \), defined as \( f^* : T \to \mathbb{R}^k \). In Meira et al. (2016), \( f^* \) represents the four production variables \( \text{NPV}, \text{ORF}, \text{Np}, \text{and} \text{Wp} \). For the sake of simplicity, assume that \( f_a(s_i) \neq f_a(s_j) \) if \( s_i \neq s_j \). In RMFinder 2.0, as in the previous version, the user is supposed to provide a set of weights \( q^r = (q_1^r, \ldots, q_k^r) \in \mathbb{R}^k_+ \), which corresponds to the subjective importance given for each risk curve of a production variable \( f_a \).

Now consider a set of representative models \( \tau' \subset T \), with \( k \) elements, and a probability distribution \( P = (p(1), \ldots, p(k)) \), where \( p(r) \) represents the individual probability \( p(r, \tau') \) of representative model \( r \) (since \( \tau' \) is clear in this context). Each representative model \( r \) also has a cumulative probability \( p_c(r, \tau') \), given by Eq. (2). Therefore, if \( \Psi \) is the algorithm that calculates \( P \), then:

\[
P \leftarrow \Psi(T, \tau', f^*, q).
\] (7)

In other words, the probability distribution \( P \) is obtained from the original set of scenarios \( T' \), the set of RMs \( \tau' \), the production variables \( f^* = (f_1, \ldots, f_k) \) and the set of weights \( q \).

Given that the set of RMs is supposed to properly represent the original set of scenarios \( T' \), it is clear that each RM \( r \) represents a subset of scenarios from \( T \). Therefore, it is possible to define \( rm^f(s) \) (Eq. (8)), which indicates by which RM \( r \) each scenario \( s \) of \( T \) is represented, given a production variable \( f \in (f_1, \ldots, f_k) \).

\[
rm^f(s) = \left\{ r \mid r \in \tau' \text{ and } p_c(r, \tau') - \frac{p_c(r, \tau')}{2} \leq p_c(s, T) < p_c(r, \tau') + \frac{p_c(r, \tau')}{2} \right\}.
\] (8)

The mechanisms behind the calculation of \( rm^f(s) \) are illustrated in Figure 6a for NPV, where the purple dots represent the original scenarios in \( T' \), the black dots the RMs and the size of the vertical bars indicate the
individual probability of each RM. If all the probabilities involved in this example are considered, it is possible to see that scenario \( s \in T \) is represented by \( r \in r^* \)(s is covered by the vertical bar of \( r \)).

\[
\Delta f'(P) = \frac{1}{n} \sum_{s \in T} |f(s) - f(rm_f'(s))|.
\]

The rationale behind the procedure described in Eq. (9) is illustrated in Figure 6b (for the NPV risk curve): \( \Delta f'(P) \) is given by the sum of the lengths of the horizontal black lines associated to each \( s \in T \), multiplied by the probability \( 1/n \). As in the original version of RMFinder, the values of the production variables \( f^* = (f_1, \cdots, f_4) \) must be normalized, so that the distance between the first and third quartiles becomes 1.0.

However, the goal is to obtain the probabilities \( P \) that minimize, at the same time, \( \Delta f'(P) \) for all production variables \( f \). Therefore, if we consider the four original production variables adopted by RMFinder, the optimization problem that must be solved so that \( P \) can be obtained becomes:

\[
\min \sum_{i=1}^{4} q_i^r \Delta f_i(P)
\]

\[
\text{s.t.} \sum_{i=1}^{k} p_i = 1, \quad p_i \geq 0, \quad i \in \{1, \ldots, k\},
\]

where \( k \) is the number of representative models and \( q_i^r \) is the weight defined by the user for the risk curve of the production variable \( i \).

Assuming that \( P^* \) is the set of probabilities that minimizes the problem defined in Formulation (10), then \( P^* \) is the ideal set of probabilities to be adopted by RMFinder 2.0 for a given set of RMs. In practice, a high quality solution \( P^* \) is assigned to a given set of RMs.

**Important aspects about the algorithm to obtain the probabilities**

It is not difficult to solve the optimization problem defined in Formulation (10) to obtain \( P^* \), as it has a convex domain. However, there is a complicating factor that should not be ignored: RMFinder 2.0 uses a metaheuristic to obtain a high quality set of RMs. Such strategy can easily evaluate more than 500,000 sets of RMs. Given that a high quality solution \( P^* \) must be obtained for each set of RMs, if the optimization algorithm that searches for \( P^* \) spends 1 second, almost 139 hours would be added to the total optimization time required to obtain the RMs. Therefore, it is imperative that \( P^* \) is obtained by an efficient algorithm.

In our implementation of RMFinder 2.0, an iterative greedy optimization algorithm was developed to search for \( P^* \) for each set of RMs. It was verified that such algorithm is capable of obtaining a solution \( P^* \) that approximates \( P^* \) with a computational time of \( O(n \cdot \log(n) + k \cdot n + k^3) \), where \( n \) is the total number of scenarios \( s \in T \) and \( k \) is the desired number of representative models. In practice, this algorithm was capable of obtaining a high quality solution \( P^* \) for a given set of RMs within milliseconds, which adds about 500 seconds to the total optimization time.
of RMFinder 2.0, considering an execution in which 500,000 different sets of RMs are evaluated during the search.

However, it is important to highlight that any optimization algorithm can be employed to solve the problem given in Formulation (10), as long as its computational time does not significantly impact the overall optimization process that is performed to obtain a high quality set of RMs.

The new objective function of RMFinder 2.0

Besides the addition of the capabilities to obtain the probability of each RM in $\tau'$, the objective function that is supposed to be minimized by RMFinder was also improved (Eq. (11)).

$$ F(\tau') = q_1 \cdot F_{\text{cross}}(\tau') + q_2 \cdot F'_{\text{risk}}(\tau') + q_3 \cdot F'_{\text{attr}}(\tau') + q_4 \cdot F_{\text{freq}}(\tau'), $$

(11)

where $F_{\text{cross}}(\tau')$ is the cross-plot component, given by Eq. (4), $F_{\text{freq}}(\tau')$ returns the number of attribute levels of the problem that are not present in the set of RMs, $F'_{\text{risk}}(\tau')$ is the new risk-curve component and $F'_{\text{attr}}(\tau')$ is the new attribute-level component of the objective function. The weights $q_i, i \in \{1, ..., 4\}$, must be defined by the user, so his/her preferences (importance of each component) can be reflected on the objective function.

The new risk-curve component, $F'_{\text{risk}}(\tau')$ given in Eq. (12), was introduced for the sake of consistency, so that a good representativeness of the risk curves can be achieved when the differences between the risk curves are low. Therefore, the same distance function given in Eq. (9) was adopted here, which is based on the set $P$ of probabilities of the RMs in $\tau'$.

$$ F'_{\text{risk}}(\tau') = \min_{\Delta \varphi \in P} \left\{ \sum_{i=1}^4 q_i^* \Delta i^*(P) \right\} $$

(12)

The weights $q_i^*, i \in \{1, ..., 4\}$, indicate the preferences of the user with respect to the representativeness on the individual risk curves for each production variable.

The last modified component of RMFinder 2.0’s objective function was $F'_{\text{attr}}(\tau')$, given in Eq. (13).

$$ F'_{\text{attr}}(\tau') = F'_{\text{dist}}(\tau'), $$

(13)

where $F'_{\text{dist}}(\tau')$ evaluates the difference between the distribution of attribute-levels in $\tau'$ and in the original set of scenarios. This component is similar to the one adopted in the original RMFinder, but the original penalization term was replaced by the fourth component of the objective function given in Eq. (11), $F_{\text{freq}}$.

Finally, the four components of the objective function given in Eq. (11) must be normalized, so that each one presents values within the same interval. In this work, the same normalization process described in Meira et al. (2016) was adopted.

The new output of RMFinder 2.0

Another improvement added to RMFinder 2.0 was the new graphical report that is generated at the end of each run. The goal was to generate a report of the obtained results so that the user can easily evaluate their overall quality (representativeness of the RMs within the original set of models) and compare two (or more) different solutions generated by RMFinder. To achieve such goal we have decided that relevant information about the results should be presented in a single screen or page and, whenever possible, graphically. Therefore, several information visualization approaches were explored here.

An example of the proposed graphical report that is generated by RMFinder 2.0 is given in Figure 7. In this report, the user can easily evaluate the distribution of the RMs (larger black dots) over the original data of the problem (purple dots) in the cross-plots, compare the risk curves, evaluate the differences between the distribution of attribute-levels, identify the models predefined by the user in all charts (black squares), see the final value of RMFinder’s objective function associated with that solution and even compare the quality of that solution with the best solution identified by RMFinder 2.0$^2$ in that run.

Since RMFinder 2.0 was implemented considering four production variables ($Np, Wp, ORF$ and $NPV$), there are six cross-plots and four risk curves to be analyzed by the user, besides the distribution of attribute-levels (that depends on the problem being considered).

In the proposed report, the six cross-plots of the problem ($Wp \times Np, Np \times ORF, Np \times NPV, Wp \times ORF, Wp \times NPV$ and $ORF \times NPV$) are presented as a modified scatterplot matrix, which does not have the conventional

---

2 As the original RMFinder, RMFinder 2.0 is also capable of returning multiple high-quality solutions in a single run.
matrix order (see “Cross-Plots” in Figure 7). Note that the original scatterplot matrix technique would provide 16 cross-plots, which would require a bigger space on the page and would compete with the space needed by other graphics.

With respect to the risk curves, they are presented in four charts, one for each production variable. In each chart, the risk curves generated by the RMs (and their probabilities) are superposed to the ones generated by the original set of models (see “Risk Curves” in Figure 7). The RMs are also projected in the vertical axis.

The differences among the distribution of attribute-levels between the RMs and original models are also presented graphically, through a set of bar charts that indicate, in blue, the differences in frequency observed for each level (column) of each attribute (row – see “Attribute-level Differences” in Figure 7). Therefore, the “bluer” each rectangle is, the higher the frequency difference between the set of RMs and the original set of models for that particular attribute-level. RMFinder 2.0 identifies the number of attribute-levels of the problem from the dataset provided as input, and automatically generates a set of bar charts that suits such problem. Therefore, the set of bar charts indicates at the same time a macro information (the number of levels per attribute) and a micro information (the value of each difference in frequency), which characterizes the use of a micro/macro design. Besides, the values of the objective function components $F'_{atr}$ and $F'_{freq}$ are also shown in the final report, both numerically and at a bar chart (see $Sum^2$ and $Pen$, respectively, at “Attribute-level Differences” in Figure 7).

This graphic report has links to the reports generated for the other solutions obtained by RMFinder 2.0 and also for solutions predefined by the user at the beginning of the search process (see “Link to other solutions” in Figure 7). A fast graphic rendering (provided by PDF viewers) and the consistency among all reports in the same document enable users to switch among reports and to focus on the differences among them.

Finally, each report also provides numerical information that indicates how well that solution performs when compared with the best solution found by RMFinder 2.0 at that run. Such information, which is available for each cross-plot, risk curve, $F_{cross}$, $F'_{risk}$, $F'_{atr}$ and the main objective function, is a value in $[0, 1]$ that indicates the relative component value of the solution being analyzed divided by the corresponding value of the best obtained solution. For example, if the report for a given solution indicates $F_{cross} = 0.98$, this means that, for the cross-plot component of the objective function, $F_{cross}$, we have $F_{current} / F^*_{cross} = 0.98$. Therefore, such solution is 2% better than the best solution obtained by RMFinder 2.0.
The example given in Figure 7 corresponds to the best solution found by RMFinder 2.0 for the problem being considered, so all these numerical information have value 1.0. Observe that the cross-plots are grouped by soft blue background, while the risk curves have soft green background.

**Results**

RMFinder 2.0 was applied to four different problems: (i) a small synthetic problem; (ii) UNISIM-I-D, which is a benchmark case based on the Brazilian Namorado field; (iii) UNISIM-II-D<sup>5</sup>, also a benchmark case, but based on a highly fractured pre-salt carbonate reservoir; and (iv) ST001a, a highly heterogeneous heavy oil offshore field. All the results that will be presented and discussed in this section were obtained within the context of the 12-step methodology of Schiozer et al. (2015), but the focus will be given solely on the task of RM selection (Step 8 of the methodology). All computational times reported here were obtained in a computer with an Intel Core i7 2.9 GHz processor, 8GB of RAM and macOS Sierra.

**Synthetic Model**

The first application of RMFinder 2.0 was to a small synthetic model with 4 (four) uncertain attributes, all discretized into 3 different levels. These uncertain attributes lead to 3<sup>4</sup> = 81 different scenarios, which were all considered for the selection of RMs.

The goal of the experiments made with this synthetic model was to evaluate the impact of the number of RMs on the representativeness of the original set, specifically in the risk curves. Therefore, three sets of RMs, with different sizes, were obtained: RM<sub>1</sub> with 3 models, RM<sub>2</sub> with 5 models and RM<sub>3</sub> with 9 models. Besides, RMFinder’s functionality that allows the pre-definition of a few RMs in the RM set was explored here, so that RM<sub>1</sub> ⊂ RM<sub>2</sub> ⊂ RM<sub>3</sub>. The weights adopted in the experiments for each component of RMFinder 2.0’s objective function are given in Table 1.

<table>
<thead>
<tr>
<th>Main Components</th>
<th>Cross-plot Components</th>
<th>Risk-curve Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>q&lt;sub&gt;1&lt;/sub&gt;</td>
<td>q&lt;sub&gt;2&lt;/sub&gt;</td>
<td>q&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>3.0</td>
<td>6.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 1 – Weights adopted for RMFinder 2.0’s objective function (Synthetic Problem).

Considering that all 81 scenarios of this problem were labeled with IDs from 0 to 80, the IDs of the RMs obtained by RMFinder 2.0 for RM<sub>1</sub>, RM<sub>2</sub> and RM<sub>3</sub> and their individual probabilities were:

- RM<sub>1</sub> = {17 30 73};
  Probabilities<sup><sub>RM<sub>1</sub></sub></sup> = {0.247 0.333 0.420};
- RM<sub>2</sub> = {17 30 73 3 46};
  Probabilities<sup><sub>RM<sub>2</sub></sub></sup> = {0.086 0.259 0.296 0.161 0.198};
- RM<sub>3</sub> = {17 30 73 3 46 22 32 55 71};
  Probabilities<sup><sub>RM<sub>3</sub></sub></sup> = {0.086 0.086 0.149 0.136 0.136 0.111 0.086 0.099 0.111}.

Graphically, each set of RMs obtained for this problem can be seen in figures 8, 9 and 10, which correspond to the standard output developed for RMFinder 2.0. As expected, an increase in the number of RMs leads to a better representation of the original set of models. This better representation can be visually noticed in the cross-plots and, particularly, in the risk curves shown in those figures.

Differently from the individual values of each production variable, the other models in the set directly affect the cumulative probability of a particular RM, which has a severe impact on the risk curves. As it is possible to see in figures 8, 9 and 10, the risk curves obtained for larger sets of RMs are closer to the original risk curves of the problem, thus better representing them.

Besides the good distribution observed in the cross-plots and risk curves, RMFinder 2.0 also tries to keep the distribution of attribute levels in the sets of RMs as close as possible to that observed in the original set. For RM<sub>1</sub>, it is possible to see in Table 2 that the exact same distribution was obtained. Table 2 highlights the differences between the attribute level distributions, both numerically and graphically, and also indicates that, for RM<sub>1</sub>, all attribute levels were represented in the RM set (Penalty = 0). The differences between the distribution of attribute levels are also graphically given in figures 8-10 (at the top-right corners), from which is possible to see that the results obtained for RM<sub>2</sub> and RM<sub>3</sub> were also satisfactory.
Figure 8 – RMFinder’s output with the cross-plots, risk curves and attribute-level distribution for RM₁.

Figure 9 – RMFinder’s output with the cross-plots, risk curves and attribute-level distribution for RM₂.
Das probabilidades gastou 0.38ms por conjunto, totalizando 2m50s no total. O algoritmo genético testa 446mil conjunto de modelos representativos por execução. O calculo foi utilizado para calcular as probabilidades dos modelos propostos em UNISIM-I-D. Neste trabalho não foi levado em conta as probabilidades dos MRs. O RMFinder 4.2 UNISIM-I-D

Table 1: Relative frequencies for attribute-level combinations considering the original set of scenarios

<table>
<thead>
<tr>
<th>Atrib.</th>
<th>set = T'</th>
<th>set = R_00</th>
<th>Difference</th>
<th>sum^2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lev. (0, 1, 2, 3*, 4*)</td>
<td>lev. (0, 1, 2, 3*, 4*)</td>
<td>lev. (0, 1, 2, 3*, 4*)</td>
<td></td>
</tr>
<tr>
<td>So</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.00,0.00,0.00,0.00)</td>
<td>0.00</td>
</tr>
<tr>
<td>Krw</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.00,0.00,0.00,0.00)</td>
<td>0.00</td>
</tr>
<tr>
<td>Por</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.00,0.00,0.00,0.00)</td>
<td>0.00</td>
</tr>
<tr>
<td>Stc</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.33,0.33,0.33)</td>
<td>(0.00,0.00,0.00,0.00)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Freq. Zero Num. (Penalty): 0       Total: 0.00

Table 2 – Relative frequencies for attribute-level combinations, considering the original set of scenarios and RM_1 (labeled R_00).

Figure 10 – RMFinder’s output with the cross-plots, risk curves and attribute-level distribution for RM_3.
Table 2 shows the same data visualized in two distinct ways: numerically and graphically. This illustrates that graphical visualizations may help the user to comprehend the information provided faster than when textual results are provided. This is important in RMFinder 2.0’s context, since it is an interactive tool.

In order to obtain RM1, the meta-heuristic adopted here (cob-aiNet[C]) evaluated 20,000 sets of RMs and demanded a total time of 6.0 s. To obtain the individual probabilities of the elements in a single set of RMs, the proposed approach demanded 0.15ms (3s in total), which is about 50% of the total time required to obtain the RMs. Similar results were observed for RM1 and RM2, for which the proposed approach required 0.07ms and 0.11ms to calculate the individual probabilities for the elements in a single set of RMs, respectively. The fact that 3 models were predefined in RM2 and 5 models in RM3 makes the optimization step much easier, since many unfeasible sets of RMs are discarded.

**UNISIM-I-D: A Benchmark Model based on the Namorado field**

The first experiment made here with a real-world-based model was with the model known as UNISIM-I-D (Gaspar et al., 2015). Meira et al. (2016) already studied UNISIM-I-D, but the individual probabilities of the RMs were not considered in their work. This benchmark problem, based on the Brazilian field named Namorado, has 11 uncertain attributes with different numbers of levels (ranging from 2 to 5), which results in more than 500,000 possible scenarios. However, in order to keep this problem feasible, such large number of scenarios has been previously filtered in the initial steps of the 12-step methodology of Schiozer et al. (2015). Therefore, at Step 8, in which RMFinder 2.0 is supposed to be applied, the original set of scenarios was reduced to 214 models (all identified by IDs that range from 0 to 213).

Besides, decision-makers that are expert on the definition of production strategies demanded that the models with IDs 83 and 144 be present on the RM set, as model 83 is the closest one to the P50 model (which has cumulative probability close to 50%) and model 144 is the base model indicated by geologists. The number of RMs set is 9. The weights adopted in this experiment for each component of the objective function are given in Table 3.

![Table 3](image)

The solution reported by Meira et al. (2016) was manually added to the RMFinder 2.0, which obtained the following probabilities:

- \( RM_{UNISIM-I-D} = \{182, 157, 185, 38, 83, 1, 10, 144, 90\}; \)
- \( Probabilities^{UNISIM-I-D} = \{0.042, 0.103, 0.121, 0.131, 0.173, 0.145, 0.140, 0.075, 0.070\}; \)

The representativeness of the RMs with respect to the cross-plots, risk curves and attribute-level distribution can be seen in Figure 11. In Figure 11, the predefined models are represented by square dots. These results were evaluated by the decision makers involved with UNISIM-I-D and considered suitable, so \( RM_{UNISIM-I-D} \) was adopted as the standard set of RMs in the following steps of the 12-step methodology of Schiozer et al. (2015).

In order to obtain \( RM_{UNISIM-I-D} \), the meta-heuristic implemented in RMFinder 2.0 evaluated 446,000 sets of RMs, which required 4 minutes and 27 seconds of processing time. To calculate the probabilities of each RM set of UNISIM-I-D, RMFinder 2.0 required 0.38ms, so the overall computational time spent with probabilities calculation was 2 minutes and 50 seconds, or 64% of the total time. Although this percentage can be considered high within the general process, the total time to automatically obtain the set of RMs is not significant.

To evaluate whether RMFinder 2.0 can obtain high quality sets of RMs in feasible time, another experiment was made, in which 30 RMs were supposed to be identified by the algorithm. The 9 RMs previously obtained \( RM_{UNISIM-I-D} \) were predefined in the initial configuration, so the task of the algorithm was to identify 21 new models that, together with the predefined ones, properly represent the original 214 models of the problem. The obtained results are shown in Figure 12, and the same set of weights given in Table 3 was adopted here.

Figure 12 shows the best set of 30 RMs obtained by RMFinder 2.0 properly represented the original set of models in the cross-plots, attribute-level distribution and, specially, risk curves. The results given in Figure 12 also show, when compared to those reported in Figure 11, that a larger number of RMs leads to a better representativeness of the original set, which is consistent with what was observed for the Synthetic Problem.
Table 4. The IDs of the RMs obtained by RMFinder 2.0, together with their individual probabilities, are:

- \( R_M^{\text{UNISIM-II-D}} = \{174, 248, 161, 285, 188, 176, 164, 163, 68\} \)
- \( \text{Probabilities}^{\text{UNISIM-II-D}} = \{0.096, 0.099, 0.077, 0.141, 0.141, 0.128, 0.109, 0.119, 0.090\} \)

With respect to the computational times, RMFinder 2.0 required 74m23s to obtain the set of RMs presented in Figure 12, since the implemented meta-heuristic evaluated 2.6 million possible sets of RMs. The average time to obtain the probabilities for each set of RMs was 1.1ms per set, which led to a total time of 48 minutes, or about 65% of the total time required by RMFinder 2.0. This is consistent with the experiment performed with 9 RMs. As it is possible to see, the total time required by RMFinder 2.0 increases significantly with the number of RMs, but such increase can be mitigated with a different adjustment of the meta-heuristic parameters. In the experiments performed here, the values recommended by Meira et al. (2016) were adopted for such parameters.

Although the computational times are higher, it is possible to see that RMFinder 2.0 is capable of dealing with different numbers of RMs, always trying to obtain a good representativeness with respect to the cross-plots, risk curves and attribute-level distribution. Therefore, it can be used to obtain sets of RMs with more than the traditional 9 or 10 models that are usually adopted in methodologies focused on production strategy definition, such as Step 8 of the 12-step approach of Schiozer et al. (2015). This may enrich the decision-making process and even open up possibilities of new methodologies to further develop oil fields.

UNISIM-II-D

UNISIM-II-D is also a benchmark case from the literature, but based on a highly fractured pre-salt carbonate reservoir (Correia et al., 2015). As this case is newer and more complex than UNISIM-I-D, reservoir experts are still developing its production strategy according to the 12-step methodology of Schiozer et al. (2015). Therefore, the RMs that will be presented here were obtained from a preliminary set of 312 models, so they will be referred here as UNISIM-II-D.

This benchmark case has 10 uncertain attributes, with different number of levels. From all possible models, reservoir experts filtered 312 feasible ones, which were labeled with IDs that range from 0 to 311. In this experiment, RMFinder 2.0 was configured to obtain 9 RMs, with the weights of the objective function given in Table 4. No models were predefined in the RM set.

The IDs of the RMs obtained by RMFinder 2.0, together with their individual probabilities, are:

- \( R_M^{\text{UNISIM-II-D}} = \{174, 248, 161, 285, 188, 176, 164, 163, 68\} \)
- \( \text{Probabilities}^{\text{UNISIM-II-D}} = \{0.096, 0.099, 0.077, 0.141, 0.141, 0.128, 0.109, 0.119, 0.090\} \)
RMFinder 2.0 was applied to obtain a set of RMs. The last problem that was considered in this work was ST001a, which required 13 levels of the original set were represented in the RMs, as the original set of 312 models was high heterogeneous. The weights define the representativeness of the RMs with respect to the cross-plots, risk curves and attribute-level distribution.

Table 4 – Weights adopted for RMFinder 2.0's objective function (UNISIM-II-D^β).

The representativeness of the RMs with respect to the cross-plots, risk curves and attribute-level distribution can be seen in Figure 13. These results indicate that the set of RMs obtained by RMFinder 2.0 properly represent the original set of 312 models according to all criteria considered here. It is also possible to see that all attribute levels of the original set were represented in the RMs, as Pen = 0 (Figure 13).

To obtain the set of RMs for UNISIM-II-D^β, RMFinder 2.0 evaluated almost 889,000 different sets of RMs, which required 13 minutes and 2 seconds of processing time. To calculate the probabilities of each RM set of UNISIM-II-D^β, 0.59ms were required, so the overall computational time spent with probabilities calculation was 8 minutes and 44 seconds, or 67% of the total time. Such results are similar to what was observed for UNISIM-I-D.

ST001a
The last problem that was considered in this work is ST001a, which corresponds to a highly heterogeneous heavy oil offshore field. ST001a has three uncertain attributes, with 5, 5 and 7 possible values each, which leads to 5 · 5 · 7 = 175 possible scenarios. From these full set, 100 feasible scenarios were filtered by experts, and RMFinder 2.0 was applied to obtain a set of 9 RMs. Again, no models were predefined in the RM set.

The weights defined for the objective function are given in Table 5, and the IDs of the RMs in the best set, together with their probabilities were:

- \( R_{M_{\text{ST001a}}} = \{44, 8, 83, 31, 72, 4, 22, 59, 20\} \);
- \( \text{Probabilities}_{\text{ST001a}} = \{0.059, 0.109, 0.139, 0.119, 0.149, 0.188, 0.129, 0.049, 0.059\} \);
With respect to the computational times, RMFinder 2.0 evaluated almost 327,000 different sets of RMs during the optimization process, which required 94 seconds of processing time. To calculate the probabilities of each RM, 0.18 ms were required, so the overall computational time spent with probabilities calculation was about 59 seconds, or 63% of the total time.

### Conclusions

In this paper, an improvement of the RMFinder technique, originally proposed to automatically and iteratively identify sets of Representative Models within decision-making procedures for oil fields, was proposed. In this new version, named RMFinder 2.0, a new approach to calculate the individual probabilities of each RM was developed, together with improvements in the objective function that is optimized by RMFinder and a new graphical report that eases the visualization and comparison of results by the user.
RMFinder 2.0 can be used to obtain sets of RMs with more than the traditional 9 or 10 models that are usually adopted in methodologies focused on production strategy definition. The UNISIM-I-D results show that the representativeness increases with the number of RMs. See figures 11 and 12 for RMs sets of size 9 and 30, respectively.

RMFinder 2.0 was applied to four different problems, being one synthetic and three based on real reservoirs. The obtained results indicated that RMFinder is capable of obtaining sets of RMs that properly represent the original set of models of the problem with respect to the cross-plots, risk curves and distribution of attribute levels.

The results observed for the risk curves are of particular interest in this work, as the mechanisms that, differently from the original RMFinder, allow the RMs to have distinct probabilities. It was shown in this work that such (possibly) different probabilities allow risk curves generated by the set of RMs to be much closer to the original risk curves of the problem, and this was confirmed in the experimental results.

With respect to the computational times required by RMFinder 2.0 to obtain the sets of RMs, the inclusion of the module to calculate the individual probabilities led to an average increase of about 65% of the total processing time. However, no more that 15 minutes were spent in each run of RMFinder 2.0 in the experiments reported here, which indicates that RMFinder 2.0 is still a feasible tool to be used in the decision making process to define production strategies in oil fields, particularly if it is considered that previous manual approaches often required much more.

Therefore, it is possible to conclude that RMFinder 2.0 is a relevant tool to aid the development of production strategies in oil fields. Such tool even allows the improvement and development of new decision-making techniques, as it automatizes an important step within such techniques.

Acknowledgements

This work was carried out with the support of the project “Methodologies for Oil Field Development and Management through Reservoir Simulations” (UNICAMP/PETROBRAS-CENPES/ANP) and under ANP R&D “Commitment to Investments on Research and Development”. The authors would like to thank PETROBRAS and Foundation CMG for the financial support and CMG and Schlumberger for the software licenses.


References


