USING REORDERABLE MATRICES TO COMPARE RISK CURVES OF REPRESENTATIVE MODELS IN OIL RESERVOIR DEVELOPMENT AND MANAGEMENT ACTIVITIES

Celmar Guimarães da Silva¹, Luis Augusto Angelotti Meira¹, Antonio Alberto S. Santos² and Denis José Schiozer³

¹School of Technology, University of Campinas, Limeira, SP, Brazil
²Center for Petroleum Studies, University of Campinas, Campinas, SP, Brazil
³School of Mechanical Engineering, University of Campinas, Campinas, SP, Brazil

ABSTRACT
Methodologies of oil reservoir development and management demand the creation of a set of reservoir models that represents the uncertainties of a reservoir. This set of uncertainties is often simplified by the use of Representative Models (RMs), i.e., models that represent the full set. Comparison of risk curves (a.k.a. complementary cumulative distribution functions) of reservoir variables is an approach used for helping oil engineers to select a set of RMs. A typical comparison chart of a given variable of interest superposes two risk curves: one of the entire model set, and another from the set of RMs. The level of similarity between the curves in this chart indicates the representativeness of the set of RMs regarding the entire model set. A visualization with some of these charts may help to compare the representativeness of the set of RMs regarding more variables (reservoir properties, production data etc.) or distinct sets of RMs. However, this kind of chart is not enough to provide an overview of these comparisons if the number of variables or sets of RMs increase. This paper shows how the use of reorderable matrices, depicted as heatmaps, can provide an overview of this dataset that can be helpful to engineers to make decisions. We propose to represent the dissimilarity of pairs of risk curves instead of the curves themselves. This solution enables our visualization to increase the number of sets of RMs and the number of variables to represent. We show the usefulness of our proposal in three case studies of oil reservoir benchmarks, and discuss the pattern we found in these cases.

KEYWORDS
Reorderable Matrix, Heatmaps, Oil Reservoir Model, Cumulative Distribution Functions, Risk Curves, Representative Models

1. INTRODUCTION
Oil reservoir development and management activities include, among others, a set of tasks related to reservoir characterization, uncertainties reduction, production forecast, and risk analysis. A reservoir structure is often modeled as a set of tridimensional grid blocks. Petrophysical properties like porosity and permeability of a given reservoir grid block, and the initial amount of oil, gas, and water inside this block are some of the other data that the model tries to estimate. The uncertain behavior of the reservoir along the time is also modeled, according to a set of possible exploitation strategies. The overall optimization process is guided by the analysis of production and economic data (output variables), such as the net production value and the volume of oil and gas produced along time. Examples of decision analysis to be taken by oil enterprises are where and when to place, open and close wells, in order to maximize their profit. All these variables have associated uncertainties, and therefore many models and strategies must be created, simulated, and analyzed before taking a decision (Schiozer et al., 2019).
An interactive selection of representative models (Schiozer et al., 2019) may help to reduce the overall computational time spend when selecting production strategies. A representative model (RM) represents a subset of the entire set of models created for a reservoir (Mahjour et al., 2020). A good set of RMs (also referred as solution in this paper) must faithfully represent the entire set of models.

The software RMFinder (Meira et al., 2016, Meira et al., 2020) uses optimization and visualization procedures to help decision makers to find good solutions. Part of these solutions is a set of risk curve charts, one for each variable. A risk curve chart contains two risk curves (a.k.a. complementary cumulative distribution functions or CCDFs) of a variable: one for the entire set of models, and other for the RMs only (see Section 3 for details). RMFinder returns a set \( S \) of best solutions to be visually analyzed by decision makers.

A problem arises when the set of variables or the number of solutions increases. To compare \( |S| \) best solutions with \( |V| \) variables each one, the decision maker must analyze \( |S| \times |V| \) risk curve charts. If the decision makers want to analyze all \( |V| \) variables and not only a subset of them, this number of charts makes unfeasible to do a reliable visual analysis of the results to find the best solution among the ones selected by RMFinder.

This paper tries to overcome this difficulty regarding risk curve charts by providing an overview of their data. Given that the main information of each of these charts is the difference of its pair of risk curves, we calculate the area \( a(s,v) \) between both curves of a chart as an error, and use this measure as an indicator of the quality of the solution \( s \), for the variable \( v \). We construct a reorderable matrix with \( |S| \times |V| \) cells, where each cell has the value of \( a(s,v) \), and depict this matrix as a heatmap. Matrix reordering algorithms help to identify patterns in the set of solutions, and the behavior of these solutions for each variable. The novelty of our approach is the visual data summarization that we provide to analysts as an overview of the error measures of many pairs of risk curves.

We applied our approach to RMs from three synthetic oil reservoirs. In our analysis, the reordering algorithm unveiled an interesting order of the variables. For each solution, variables at the end of this order have tendentially more error in a given solution than the variables at the start of the order, regardless of the reservoir analyzed.

This paper is organized as follows. Section 2 briefly introduces theoretical background related to reservoir modeling, representative models, and reorderable matrices. Section 3 summarizes visualization strategies to compare risk curves. Our visualization proposal is defined at Section 4. Section 5 shows how our reorderable matrices present data from synthetic oil reservoir benchmarks. Section 6 concludes our paper and points out future work.

2. THEORETICAL BACKGROUND

In this section we briefly introduce concepts about reservoir modeling in oil engineering research. Besides, we present concepts related to reorderable matrices, which we use in our approach.

2.1 Reservoir Modeling Concepts

In the context of this work, a reservoir model can be defined as a tridimensional grid, in which each cell (grid block) has values of a set of attributes such as oil, gas and water saturation, porosity, and permeability, among others. In other words, a model \( m \) can be defined as a set of tuples in the format \([i, j, k, iv, val]\), where \( i, j \) and \( k \) correspond to the axes X, Y, and Z; \( iv \) is an input variable, and \( val \) is the value of this variable in the given cell (Aziz, 1993). This is a simplified way to model a reservoir; other approaches add spatial distortion to the cubic cells of this model to better represent reservoirs in the space (Correia et al., 2015; Mahjour et al., 2019).

A production strategy add wells to a reservoir model. There are two types of wells: producer wells, which extracts fluid; and injector wells, which insert fluid into the reservoir to help oil and gas extraction by increasing the internal pressure inside the reservoir (Dake, 1983).

Given a model and a production strategy, a reservoir can be simulated in a specialized software. This kind of simulation generates two types of output variables: field variables, such as water production (Wp), net production (Np), net present value (NPV), oil recovery factor (ORF), oil in place (OIP), water in place...
(WIP), and water injection (Winj). It also outputs well variables, such as well economic index (WEI), among others.

As presented in the introduction, the selection of RMs is a relevant step that makes it feasible to run simulations based on models. Choosing the best RMs for a given set of models is not straightforward. This task may be accomplished by methodologies based on ranking, clustering or optimization (Meira et al., 2020). In this paper, we focus on RMFinder’s interactive optimization methodology (Meira et al., 2020).

Given a set of $M$ reservoir models, RMFinder tries to choose a subset $R$ with $k$ models of a set $M$ of models such that the representativeness of this subset is optimal. Three aspects drive this optimization process. The first aspect is to find well-distributed models in $M$, according to all possible 2D projections of pairs of the field variables under consideration. A second aspect is that the models in $R$ should have a distribution similar to those in $M$ for each field variable, according to classes of values predefined for each variable (a.k.a. attribute-levels). In this paper, we are particularly focused on the third optimization aspect: to find $R$ that provides similar CCDFs than $M$ for each field variable.

Regarding user interaction in this methodology, decision makers can redefine models before optimization starts. After optimization, RMFinder shows cross-plots, risk curve charts, and bar charts that help decision makers to analyze the optimization results. If the optimization results are not satisfactory, the decision makers can change weights in each of three aspects of the optimization process according to their interests and run it again.

### 2.2 Reorderable Matrices

Reorderable matrices (Bertin, 2010) are data structures that store values of a function $c(x,y)$, where $x$ and $y$ are nominal variables, and $c$ is often a quantitative function. They are frequently depicted as heatmaps, where $x$ and $y$ are mapped to X and Y axes, and $c$ is mapped to colors according to a color map. Given that $x$ and $y$ are nominal, the rows and columns of the heatmap may be reordered to reveal patterns that are hidden due to unsuitable row and column permutations.

Wilkinson (2005) and Behrisch et al. (2016) indicate a set of patterns that a heatmap may present. We highlight the Simplex pattern (Wilkinson, 2005), which we observed in the context of this work. Wilkinson (2005) defines a Simplex pattern as a matrix $X_{n\times p}$ whose cell values are $x_{ij} = e^t / (1 - e^t) + u_{ij}$ where $i = 1, ..., n; j = 1, ..., p; t = (s_j - r_i); s_j = jnp; r_i = in; $ and $u_{ij}$ is a random error. In other words, when a matrix has a Simplex pattern, its values tend to decrease from northeast to southwest, as depicted in Figure 1.

A pre-Simplex matrix is a matrix that has at least one permutation in which it shows a Simplex pattern.

Matrix reordering algorithms permute matrices in order to unveil information. Some of them focus on revealing Wilkinson’s patterns. We highlight FVS (Silva et al., 2017), a simple and fast algorithm ($O(mn)$ for $m \times n$ matrices) for revealing Simplex patterns in pre-Simplex matrices. FVS sorts the matrix rows (columns) according to the order of the average values of each row (column). A state-of-the-art reordering method called Hybrid Sort (Silva, 2020) uses FVS to reorder a (possibly scrambled) matrix that was classified as pre-Simplex.

![Figure 1](image.png)  
**Figure 1.** Illustrative example of matrix with the Simplex pattern. Values from 0 to 1 are colored according to a blue-white-red scale.
3. VISUALIZATIONS OF SETS OF RISK CURVES

This section presents some strategies used by oil engineering works to represent data sets related to risk curves. First of all, we must compare the concept of risk curve against cumulative distribution functions (CDFs) or complementary cumulative distribution functions (CCDFs).

Suppose a random variable $X$ that has a probability measure $Q$. A CDF $F(X,v)$ or $F(Q,v)$ returns the probability that $X$ is less than or equal to $v$ (Gentle, 2009). A CCDF (a.k.a. tail distribution function) $F^C$ is defined as $F^C(X,v) = 1 - F(X,v)$ (Gass & Fu, 2016). Some works use the term risk curve (Kaplan & Garrick, 1981) as a synonym to CCDF (e.g. Schioz et al., 2017, Schioz et al., 2019, Meira et al., 2016).

A risk curve is commonly plotted in a line chart, whose X and Y axes represent, respectively, the variable $X$ and the values of $F(X,v)$ (for CDF) or $F^C(X,v)$ (for CCDF). Figure 2(a,b) has a synthetic example of a probability function for NPV and two related risk curves (for 500 and 10 models).

Figure 2. (a) Frequency values for a single field variable (NPV). (b) Two CCDFs of NPV, one for 500 models (solid line) and other for 10 models (dotted line). (c) Chart with 10 merged (superposed) risk curves for Well Economic Index (WEI) of a single well WLD. (d) Risk-return plot of oil production strategies

We identified four visual mapping approaches to overview and compare many risk curves in a single chart: superposition, summarization, juxtaposition, and hybrid approaches.

It is possible to do line superposition in the line chart, as in the “merged risk curves” chart (Meira et al., 2020) (Figure 2(c)). It superposes a set of 10 risk curves for a given variable in the best solutions found in a single execution of RMFinder. Despite its cluttered lines, this chart overviews the risk curves, and enables analysts to compare their similarities.

A summarization approach may use a scatterplot in which each mark represents a risk curve. The axes of this chart represent risk and return measures. Santos et al. (2020) map “Risk of NPV” to X axis, “Estimated Monetary Value (EMV)” to Y axis, and marks to production strategies, where each strategy has a risk curve for NPV. In Figure 2(d) we exemplify the same visual mapping with synthetic data for 199 strategies.

A juxtaposition approach places risk curve charts side-by-side in a matrix layout. As an example, Meira et al. (2020) presents a chart with 25 small risk curve charts in a $6 \times 5$ matrix layout. Each of these small charts shows 2 superposed risk curves: one for the entire set of models, and other for the nine representative.
models. The full chart does not define if the order of its risk charts is meaningful or not. Some similarity criteria could be used for this layout, such as the association between clustering and a Peano-Hilbert curve used by Silva et al. (2019) for drawing multiple reservoir models. Figure 3 exemplifies five juxtaposed risk curve charts created by RMFinder. Purple and black lines refer to the RMs and the entire set of models, respectively.

A hybrid approach is proposed also by Meira et al. (2020). They use juxtaposition and superposition of many risk curves in 20 risk charts for showing the best 10 solutions for a given oil field. We adopted the same approach in the Figure 4. The charts refer to 11 field variables (NPV, ORF etc.) and to productivity indexes of 9 wells (WEI-IN1, WEI-IN2 etc.). Purple and black lines are similar to those in Figure 3, except that there are 10 superposed black lines per chart, one for each solution.

Figure 3. Juxtaposed risk curve charts of five variables (generated by RMFinder). The RMs and the entire set of models are depicted as purple and black lines, respectively

Figure 4. Juxtaposed risk curve charts with superposed risk curves (generated by RMFinder). The RMs and the 10 solutions are depicted as 1 purple line and 10 black lines, respectively.

4. ERROR COMPARISON HEATMAPS

The previous section presented some approaches for visualizing sets of risk curves. In this paper, we propose to use heatmaps as a summarization approach for comparison of pairs of risk curves. As we presented before, given two risk curves of a same variable – one for representative models and other for the entire set of models – the sum of the areas between them can be considered as an error value (Figure 5). The smaller the sum, the more representative the RMs are. We aim to show these error values as the cells of our heatmap.
In the visual mapping proposed for our error comparison heatmaps, the Y axis shows solutions defined as best choices of RMs by a software (in our case, the RMFinder software). The X axis shows field or well variables. The color of each cell represents the error value related to the solution and variable to which the cell refers. The color map has a green-white-purple scale, where green is the absence of error and purple the greatest error in the dataset. The white color was configured according to the need of the analysts during the analysis process. We selected these colors in order to avoid conflict with other colors already used in the analysis of oil variables. Besides, given that variables and solutions are categorical variables, they may be permuted if necessary. Therefore, matrix reordering algorithms can be used to unveil patterns in this matrix to the users.

This matrix may be useful at least in the following tasks:

a) Comparing solutions with the same number of RMs to choose one of the solutions. Figure 6 exemplifies this use. It shows field and well variables for 10 solutions, each one with nine RMs. The matrix was sorted by FVS (Silva et al., 2017). The white value in the color map was empirically defined as 0.2. Columns with purple cells at the right of the matrix indicate that most of the solutions under analysis do not appropriately represent the variables WEIPR6 and WEIPR4 (Well Economic Indexes of producer wells PR6 and PR4).

b) Comparing solutions with distinct number of RMs to observe which variables produce more or less error. This comparison can also help to define criteria to choose an ideal number of RMs to be used. The section Results will present some matrices related to this task.

Figure 6. Error comparison matrix with 10 solutions (Y axis) and 25 variables (X axis). All solutions have 9 RMs. Cell colors may vary from green (absence of error) to purple (maximum error in the dataset)

5. RESULTS

We generated risk comparison matrices to represent error in the solutions created by RMFinder for three synthetic reservoir benchmarks freely available: UNISIM-I-D and UNISIM-II-D, available at UNISIM’s
website (http://www.unisim.cepetro.unicamp.br/benchmarks); and *Olympus*, available at ISAPP website (https://www.isapp2.com/optimization-challenge/timeline.html). The solutions vary according to the number of RMs in each solution. In addition, some matrices have 10 solutions for each possible size of the set of RMs. Our matrices may have field and well variables, according to the needs of the analysts.

Figure 7(a) shows the error comparison matrix of solutions for UNISIM-I-D. The names of the solutions have the format “nRM-0”, where *n* indicates the number of RMs in the solution. The value 0.2 is shown as the white color in the matrix. The dominant green region of this matrix indicates that solutions with 16 or more RMs may satisfy the analysts. It is worth noting that both axis of this matrix were sorted by FVS. Therefore, the average error of each solution have a strong correlation with the number of RMs in each solution – note that only solutions 22RM-0 and 23RM-0 are out of their normal order.

Variables OIPHB, WEI19, WEIP21, and WEIP08 have the worst average values in these solutions. The first one is an oil in place (OIP) variable of a part of the reservoir called “high block” (HB), which is at the east of the reservoir. The other variables are related to well economic indexes (WEI) of injector well 119 and producer wells P21 and P08, which are also at the same region. The results indicate that the solutions found could not completely capture some important petrophysical property uncertainties of their regions.

Figure 7(b) also shows similar matrices for UNISIM-II-D reservoirs, but with distinct ordering of variables (when compared to UNISIM-I-D). The green region of the matrix seems to be smaller than the one of UNISIM-I-D data, suggesting that only solutions with 17 or more RMs are acceptable. As in the UNISIM-I-D matrix, the rightmost variables of each of UNISIM-I-D’s matrices present the worst average values for the solutions in their respective matrices. The results for Olympus were very similar to those of UNISIM-II-D and were omitted in this paper.

In both matrices of Figure 7, the reordering algorithm FVS revealed a Simplex pattern. This pattern seems to match the information in the three cases, given that the order of the Y axis after reordering was almost the ascending order of the number of RMs. This makes sense in the context of RMs, because the more RMs, the best the representation of the entire model. However, we could not find until this moment a clear explanation about why the variables were ordered in such ways in each figure. It suggests that the order may be related to greater uncertainty in the regions of the wells with greater error, but only further investigation can confirm this hypothesis.

![Figure 7](image_url)

Figure 7. Error comparison matrix of solutions calculated for the following reservoirs: (a) UNISIM-I-D and (b) UNISIM-II-D. The number of RMs per solution varies from 2 to 25. Cell colors may vary from green (absence of error) to purple (maximum error in the dataset)

6. CONCLUSION

This work presented a new approach for visualizing differences between pairs of risk curves. Our proposal uses reorderable matrices to show the error of each set of RMs when compared to the entire set of models of a given reservoir, regarding probabilities of a set of field and well variables. When solutions with distinct
numbers of RMs are compared, our proposal has the potential to show which variables have errors that are harder to reduce; in other words, which variables need more RMs to be correctly represented.

Future works include a better definition of what is an acceptable error, and the implementation of this proposal as an interactive software with easy access to risk curve charts related to each variable and solution presented in the error comparison matrix.

ACKNOWLEDGEMENT

We gratefully acknowledge the support of EPIC – Energy Production Innovation Center, hosted by the University of Campinas (UNICAMP) and sponsored by Equinor Brazil and by FAPESP – São Paulo Research Foundation (grant number 2017/15736-3). We also thank the support of FAEPEX/PRP/UNICAMP (grant #222420). We acknowledge the support of ANP (Brazil’s National Agency of Oil, Natural Gas and Biofuels) through the R&D levy regulation. Acknowledgments are extended to the School of Technology (FT), Center for Petroleum Studies (CEPETRO), and School of Mechanical Engineering (FEM) at UNICAMP. We also acknowledge the funding from Energi Simulation and thank the Computer Modelling Group Ltd. (CMG) for software licenses and technical support.

REFERENCES

Schiozer, D.J. et al., 2019, Model-based decision analysis applied to petroleum field development and management. Rev. IFP Energies nouvelles 74, 46.