

SELEÇÃO DE MODELOS GEOLÓGICOS REPRESENTATIVOS: ANÁLISE MULTIOBJETIVO

Yuri S. Costa

Faculdade de Tecnologia (FT) - Universidade Estadual de Campinas (UNICAMP)
R. Paschoal Marmo, 1888 - Limeira, SP
y254390@dac.unicamp.br

Luis A. A. Meira

Faculdade de Tecnologia (FT) - Universidade Estadual de Campinas (UNICAMP)
R. Paschoal Marmo, 1888 - Limeira, SP
meiraft@unicamp.br

RESUMO

O processo de decisão no desenvolvimento e gestão de campos de petróleo envolve a análise de estratégias de produção. Para definir uma estratégia eficiente, é necessário avaliar modelos geológicos com variáveis incertas. Porém, a combinação dessas variáveis resulta em um número exponencial de modelos. Para viabilizar a definição de uma estratégia adequada, é fundamental selecionar um subconjunto reduzido de modelos representativos (MRs). Propomos uma abordagem multiobjetivo que seleciona uma fronteira de soluções não dominadas de MRs. Foram definidas duas funções objetivo, minimizadas pelo NSGA-III: a primeira mede a representatividade dos MRs em um espaço $\mathbb{R}^5$, enquanto a segunda avalia a distribuição dos MRs pelos percentis das variáveis de produção. O algoritmo foi aplicado ao benchmark UNISIM-I-D, e resultados foram avaliados de maneira cruzada com a ferramenta RMFinder. Os resultados indicam o potencial da abordagem multiobjetivo para seleção de MRs, e sugerem avaliação mais aprofundada das funções objetivo.

PALAVRAS CHAVE. Seleção de Modelos Representativos. NSGA-III. Otimização Multiobjetivo.

OMO - Otimização Multiobjetivo, EN&OG - PO na Área de Energia, Óleo e Gás

ABSTRACT

The decision-making process in oil field development and management involves analyzing production strategies. To define an efficient strategy, geological models composed by uncertain variables must be evaluated. However, combining these variables results in an exponential number of models. To make the definition of an appropriate production strategy feasible, it is essential to select a reduced subset of representative models (RMs). We propose a multiobjective approach that evaluates and selects a Pareto front of RM solutions. Two objective functions were defined and minimized by NSGA-III: the first evaluates the representativeness of the RMs in an $\mathbb{R}^5$ space, while the second assesses their distribution across the percentiles of the production variables. The algorithm was applied to the UNISIM-I-D benchmark, and results were compared with the solution proposed by the RMFinder tool. The findings indicate the potential of the multiobjective approach for RMs selection and suggest further evaluation of the objective functions.

KEYWORDS. Representative Models Selection. NSGA-III. Multiobjective Optimization.

OMO - Multiobjective Optimization, EN&OG - OR in Energy, Oil, and Gas

1. Introdução

A exploração de campos de petróleo depende de projetos custosos e de longo prazo que envolvem um conjunto de variáveis incertas que abrangem (i) características próprias do reservatório de petróleo; (ii) aspectos operacionais; e (iii) questões econômicas, como o preço do petróleo e custos associados [Schiozer, Denis José et al., 2019]. A consideração das combinações de tais incertezas é importante para que sejam tomadas decisões adequadas no gerenciamento dos campos de exploração de petróleo a fim de gerar estratégias de produção eficientes. Entretanto, todas as combinações possíveis dessas variáveis podem gerar uma quantidade exponencial de modelos a serem analisados, podendo demandar o uso de técnicas de redução [Meira et al., 2020] para possibilitar o processo de definição da estratégia de produção.

Uma estratégia de produção pode ser vista superficialmente como a infraestrutura necessária de poços, tubulações, plataformas e coletores associados a um cronograma de operação e sua localização exata no campo de petróleo [Meira et al., 2016]. A seleção e otimização de estratégias de produção está relacionada à metodologia proposta por [Schiozer, Denis José et al., 2019] para análise integrada de decisão no desenvolvimento e gestão de campos de petróleo, descrita em 12 passos:

1. Caracterizar o reservatório sob incertezas (construir modelos, desenvolver cenários e estimar probabilidades).
2. Construir e calibrar modelo de simulação por meio um Caso Base denominado “Base 0”.
3. Verificar inconsistências no Caso Base por meio de dados dinâmicos coletado de poços.
4. Gerar cenários considerando as incertezas do reservatório.
5. Reduzir o número de cenários analisando a sua viabilidade a partir de dados de produção. Dos modelos aceitos, um novo Caso Base “Base 1”, idealmente próximo ao percentil 50% (P50), pode ser selecionado para os passos seguintes se o “Base 0” falhar na validação com os dados históricos, ou ser otimista demais, ou pessimista demais. A eventual seleção do “Base 1” considera um modelo que se aproxima do P50 em todos os indicadores, representando um caso intermediário.
6. Selecionar uma estratégia de produção determinística para o Caso Base.
7. Fazer a estimativa inicial do risco da primeira estratégia de produção com todos os cenários possíveis do Passo 5.
8. Reduzir o conjunto de modelos obtidos no Passo 5 selecionando os Modelos Representativos (MRs).
9. Selecionar estratégias de produção para cada MR, como no Passo 6.
10. Selecionar estratégia de produção sob incertezas do reservatório, econômicas e outras.
11. Identificar potenciais mudanças na estratégia de produção do Passo 10.
12. Realizar análise de decisão com base em indicadores técnicos e econômicos.

O desempenho das estratégias de produção é avaliado amplamente através de simuladores. Para realizar uma simulação é necessário uma estratégia de produção aplicada a um modelo geológico determinístico, ou seja, sem incertezas.

Cada modelo geológico possui uma estratégia de produção ótima específica para ele. Tendo em vista que o processo de obter uma estratégia de produção otimizada para um único modelo requer dias de trabalho e centenas de simulações computacionais, na prática, somente um número limitado de modelos é viável [Meira et al., 2016]. Nesse sentido, faz-se necessário a aplicação do Passo 8 que utiliza a seleção de MRs para a redução do número de modelos. Os MRs são um conjunto reduzido de modelos que representam as características incertas do conjunto original obtidos no Passo 5 da metodologia de [Schiozer, Denis José et al., 2019].

Em [Schiozer et al., 2004], a seleção de MRs é baseada nas funções objetivo que se referem às variáveis Net Present Value (NPV), Cumulative Oil Production (Np), Recovery Factor (RF) e Water Production (Wp). O critério de seleção é baseado em selecionar os modelos com diferenças significativas em RF que estejam próximos dos valores de percentis P90, P50 e P10 do NPV, descrevendo uma distribuição de modelos pessimistas, prováveis e otimistas, respectivamente.

Mahjour [Mahjour et al., 2020] utiliza uma abordagem integrada de escalonamento e agrupamento multidimensional, Integrated Multidimensional Scaling and Clustering (IMC). O método apresenta uma solução estatística para a seleção de MRs que mede a semelhança entre os modelos utilizando coeficiente de correspondência simples, Simple Matching Coefficient (SMC). A metodologia inclui o uso de agrupamento hierárquico e o algoritmo K-means.

A técnica RMFinder [Meira et al., 2020, 2016] transforma o problema de seleção de MRs em uma tarefa de otimização multiobjetivo, baseada em *cross-plots* e análise de curvas de risco, além de também considerar a distribuição de probabilidade das variáveis incertas. O método, porém, faz uma soma ponderada das funções objetivo por meio de pesos escolhidos pelo usuário. Se o equilíbrio das funções objetivo não estiver de acordo com as expectativas do usuário, ele deve trocar os pesos e executar o programa novamente.

1.1. Objetivos

O objetivo principal deste trabalho é propor um método multiobjetivo para realizar a seleção de MRs descrita no Passo 8 da metodologia de [Schiozer, Denis José et al., 2019], oferecendo uma fronteira de soluções não dominadas ao final do processo. Para alcançar o objetivo principal, será necessária a realização dos seguintes objetivos específicos:

1. Propor funções objetivo para o problema. Foi proposta uma função objetivo que avalia a representatividade em um espaço $\mathbb{R}^5$ e outra relacionada distribuição dos MRs nos percentis das variáveis de produção, sendo detalhadas na Seção 2.1.1 e Seção 2.1.2, respectivamente.
2. Uma vez definida as duas funções objetivo, implementar um algoritmo genético para encontrar conjuntos de MRs otimizados.
3. Gerar a fronteira de Pareto.
4. Comparar resultados com o RMFinder.

2. Metodologia

Nesta pesquisa foi utilizado o benchmark sintético UNISIM-I-D [Gaspar et al., 2015], construído com base em dados do campo Namorado, na bacia de Campos, Brasil. Nesta pesquisa utilizamos cinco variáveis de produção: NPV, ORF, Np, Wp e Water Injection (Winj). Destas cinco, quatro foram utilizadas no trabalho de referência [Schiozer et al., 2004]: NPV, ORF, Np e Wp.

A seleção dos modelos representativos foi realizada através de uma abordagem multi-objetivo, composta por duas funções descritas na Seção 2.1.1 e na Seção 2.1.2. Foi aplicado o Non-dominated Sorting Genetic Algorithm III (NSGA-III) para lidar com os objetivos conflitantes e obter uma fronteira de soluções não dominadas.

A análise e avaliação dos resultados desta pesquisa foi feita de maneira cruzada com o RMFinder [Meira et al., 2016, 2020], conforme os seguintes passos:

1. A fronteira de soluções encontrada neste trabalho foi inserida no RMFinder, gerando um relatório que avalia cada solução com as métricas descritas na Seção 2.4.
2. O RMFinder foi utilizado para gerar suas próprias soluções, tendo a sua melhor solução aplicada às métricas utilizadas pelas funções objetivo desta pesquisa, descritas na Seção 2.1.3. As configurações utilizadas no RMFinder são explicitadas na Seção 2.4.

2.1. Descrição do Problema

O problema recebe como entrada um conjunto de modelos M . Cada elemento $m \in M$ foi considerado um ponto no espaço $\mathbb{R}^5$, sendo representado como $m = (v_1, v_2, v_3, v_4, v_5)$, onde v_i representa uma variável de produção.

Em nosso problema, uma solução $S \subset M$ é vista como um indivíduo em um algoritmo genético. Cada indivíduo S é composto por k MRs, ou seja $|S| = k$. Tal solução é representada como um vetor $S = (mr_1, mr_2, mr_3, \dots, mr_k)$, onde cada $mr_i \in M$ é um gene que corresponde a um MR específico.

O número de MRs k entre 9 ou 10 é amplamente utilizado, uma vez que costuma resultar em um tempo razoável de simulação nas etapas seguintes à seleção de modelos representativos [Meira et al., 2016].

A saída deste problema é a fronteira de Pareto $P \subset M^k$. Cada elemento $S' \in P$ é uma solução não dominada com k MRs. Note que soluções dominadas não pertencem à fronteira de Pareto P .

Observe que o problema de seleção de MRs não possui funções objetivo bem definidas. Neste trabalho, propomos duas funções objetivo cuja satisfação e eficácia para o problema devem ser avaliadas.

2.1.1. Função objetivo F_{cluster}

A primeira função objetivo é baseado na soma das distâncias euclidianas de cada modelo $m \in M$ ao $mr \in S$ mais próximo. Dada uma solução S , o custo $F'_{\text{cluster}}(S)$ pode ser calculado da seguinte maneira

$$F'_{\text{cluster}}(S) = \sum_{m \in M} d(m, S),$$

onde

$$d(m, S) = \min_{mr \in S} d_{\text{Euclid}}(m, mr), \forall m \in M.$$

Para os custos das funções objetivo estarem em uma escala próxima, o custo desta função objetivo é normalizado com base no custo médio $\bar{F}_{\text{cluster}}$ de todos os indivíduos gerados na primeira população G_0 .

$$\bar{F}_{\text{cluster}}(G_0) = \frac{\sum_{S_i \in G_0} F'_{\text{cluster}}(S_i)}{|G_0|}.$$

Portanto, o custo da função objetivo F_{cluster} é dado por

$$F_{\text{cluster}}(S) = \frac{F'_{\text{cluster}}(S)}{\bar{F}_{\text{cluster}}(G_0)}.$$

2.1.2. Função objetivo F_{distrib}

A segunda função objetivo é utilizada para buscar soluções cujas variáveis de produção estejam bem distribuídas ao longo de seus percentis, favorecendo a escolha de modelos que mantenham uma distribuição uniforme de probabilidades acumuladas, sem um viés pessimista ou otimista. Esta função é utilizada para minimizar a distância da distribuição dos percentis dos modelos representativos em relação à distribuição considerada ideal.

Em nosso problema, todos os modelos são equiprováveis. Sendo assim, suas probabilidades são

$$\text{prob}(m, M) = \frac{1}{|M|}.$$

A probabilidade acumulada de uma variável de produção v de um modelo $m \in M$ pode ser calculada da seguinte maneira

$$\text{acc}_v(m) = \frac{1}{2|M|} + \sum_{m' \in M: m'_v \geq m_v} \frac{1}{|M|}.$$

Ou seja, a probabilidade acumulada de um modelo é a metade de sua probabilidade mais a soma da probabilidade dos modelos cuja variável de produção m'_v são maiores que a do modelo em questão m_v .

Deste modo, uma probabilidade acumulada $\text{acc}_v(m) = y$ indica que, dado um modelo $m \in M$ e uma variável de produção v , existe uma probabilidade y de que o valor final da variável v seja maior ou igual a m_v . Nesse contexto, o gráfico que representa m_v em relação a $\text{acc}_v(m)$, para qualquer variável de produção v , é conhecido como uma curva de risco [Meira et al., 2016]. A Figura 1 ilustra a curva de risco para NPV.

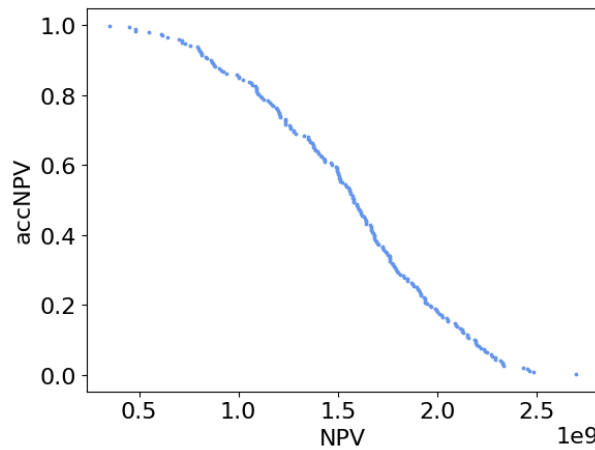


Figura 1: Curva de risco NPV $\times$ accNPV.

Outro aspecto importante desta função objetivo é definição de uma faixa de valores idealmente distribuídos, representada pelo vetor $Q = (q_1, q_2, \dots, q_k)$. Cada elemento q_i corresponde a um ponto ideal que serve como referência para a distribuição. Para k MRs, onde i assume valores no intervalo $\{1, \dots, k\}$, esses pontos q_i são definidos como valores de probabilidades acumuladas linearmente espaçadas no intervalo $[0, 1]$.

$$q_i = \frac{i}{k+1}.$$

Suponha os seguintes exemplos de como o vetor Q é definido para diferentes valores de k :

- Para $k = 1$: $Q = (\frac{1}{2})$.
- Para $k = 2$: $Q = (\frac{1}{3}, \frac{2}{3})$.
- Para $k = 9$: $Q = (10\%, 20\%, \dots, 90\%)$.

Dada uma solução S , o custo $F'_{\text{distrib}}(S)$ é calculado pela soma das diferenças quadráticas entre a probabilidade acumulada de $mr_i \in S$ e a probabilidade acumulada ideal q_i para cada variável v . Observe que o cálculo das probabilidades acumuladas é realizado de maneira independente para cada variável de produção v , com os modelos de S previamente ordenados de forma crescente em relação à v . Esse ordenamento busca garantir a correspondência direta com os valores do vetor Q .

Considere as variáveis de produção $V = (\text{NPV}, \text{NP}, \text{WP}, \text{ORF}, \text{Winj})$.

$$F'_{\text{distrib}}(S) = \sum_{v \in V} \sum_{mr_i \in S} (\text{acc}_v(mr_i) - q_i)^2.$$

Para os custos das funções objetivo estarem em uma escala próxima, o custo desta função objetivo é normalizado com base no custo médio $\bar{F}_{\text{distrib}}$ de todos os indivíduos gerados na primeira população G_0 .

$$\bar{F}_{\text{distrib}}(G_0) = \frac{\sum_{S_i \in G_0} F'_{\text{distrib}}(S_i)}{|G_0|}.$$

Portanto, o custo da função objetivo F_{distrib} é dado por

$$F_{\text{distrib}}(S) = \frac{F'_{\text{distrib}}(S)}{\bar{F}_{\text{distrib}}(G_0)}.$$

A Figura 2 exemplifica a diferença entre acc_v para uma variável de produção v de uma solução S e o vetor de distribuição ideal de percentis Q , com $k = 9$. Note que os elementos m_{v_i} estão ordenados crescentemente. A diferença entre cada m_{v_i} e q_i é representada por um segmento vermelho.

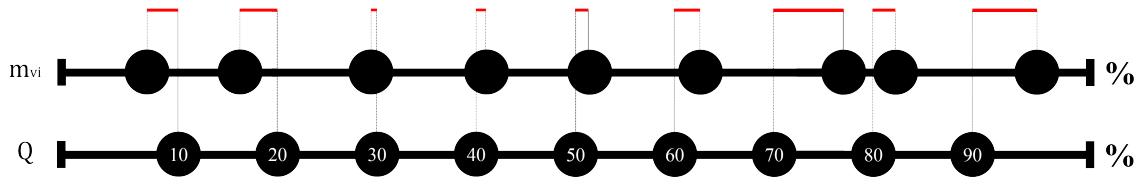


Figura 2: Representação da diferença entre m_{v_i} e q_i .

2.1.3. Cálculo do custo

A métrica de cálculo de custo para uma solução é obtida através da equação a seguir:

$$F(S) = (F_{\text{cluster}}(S), F_{\text{distrib}}(S)),$$

onde

- a) S : conjunto solução com modelos representativos.
- b) F_{cluster} : função objetivo descrita na Seção 2.1.1.
- c) F_{distrib} : função objetivo descrita na Seção 2.1.2.

Neste trabalho, o processo de estimação de custos com as funções objetivo utiliza NPV', ORF', Np', Wp' e Winj', que correspondem aos valores normalizados de NPV, ORF, Np, Wp e Winj, respectivamente. As variáveis normalizadas estão no intervalo $[0, 1]$.

2.2. Algoritmo genético

A representação de um indivíduo já foi discutida na Seção 2.1. Para a aplicação do algoritmo genético, utilizando o NSGA-III no processo de otimização multiobjetivo, foram implementados os operadores de mutação e cruzamento.

2.2.1. Operador de mutação

Na operação de mutação, dado um indivíduo $S = (mr_1, \dots, mr_k)$, uma porcentagem p de seus genes é substituída por genes aleatórios em $M \setminus S$, garantindo que os novos genes não sejam previamente existentes no indivíduo original.

A porcentagem de mutação p é determinada de forma aleatória dentro de um intervalo configurável r , onde $r \in [0, 1] \times [0, 1]$. Neste trabalho, o intervalo r foi parametrizado como $[0.25, 0.4]$, limitando a proporção de genes mutados entre 25% e 40% do total de genes do indivíduo. A escolha da posição do gene a sofrer mutação é feita de maneira aleatória.

A probabilidade de um indivíduo sofrer mutação foi fixado em 10% nos experimentos deste trabalho.

2.2.2. Operador de cruzamento

Na operação de cruzamento, dado dois indivíduos S_i e S_j , são gerados dois indivíduos filhos S_k e S_l . Este processo segue as seguintes etapas:

1. Seja $I = S_i \cap S_j$. Os indivíduos filhos S_k e S_l herdam todos os genes I .
2. Os genes $S_i \setminus I$ e $S_j \setminus I$ são alinhados. Cada par do alinhamento é dividido entre os filhos S_k e S_l de maneira aleatória.
3. Após o cruzamento, cada indivíduo filho passa por uma busca local, na qual a vizinhança é um *swap* entre dois elementos, um dentro da solução e outro fora. O processo se repete por um número fixo de iterações.

2.3. Parâmetros

Foram definidos os seguintes parâmetros para o algoritmo genético:

- Quantidade de gerações: 50.
- Tamanho da população: 60 indivíduos.
- Quantidade de indivíduos selecionados para as gerações seguintes: 30.
- Probabilidade de mutação: 10%.
- Iterações máximas na busca local: 30.

A primeira população é composta por 60 indivíduos gerados aleatoriamente. Para as gerações seguintes, a população passa a ser formada pelos 30 indivíduos de melhor avaliação da geração anterior, além de 30 novos indivíduos resultantes do operador de cruzamento aplicado a esses selecionados. Conforme exposto na Seção 2.2.2, cada par de pais gera dois indivíduos filhos.

2.4. Métricas do RMFinder

A avaliação do RMFinder [Meira et al., 2020, 2016] para mensurar representatividade é baseada em uma abordagem que considera a combinação de quatro funções objetivo, descrita pela seguinte equação

$$F_{\text{RMFinder}}(S) = q_1 \cdot F_{\text{cross}}(S) + q_2 \cdot F_{\text{risk}}(S) + q_3 \cdot F_{\text{atr}}(S) + q_4 \cdot F_{\text{freq}}(S),$$

sendo

- a) q_i : valor de peso para equilibrar a importância de cada função objetivo.
- b) F_{cross} : função objetivo que busca garantir que os modelos representativos se distribuam bem nos *cross-plots* de variáveis. São realizadas projeções espaciais em pares para quatro variáveis (NPV, Np, ORF e Wp), onde v_a e v_b representam variáveis escolhidas desse conjunto. O custo é calculado através da distorção do erro quadrático para o *cross-plot* $v_a \times v_b$, que é basicamente a soma das menores distâncias Euclidianas quadráticas entre cada modelo e o MR mais próximo.
- c) F_{risk} : função objetivo que busca garantir que os modelos representativos representem bem as curvas de risco originais para as mesmas variáveis citadas anteriormente. Originalmente, as primeiras versões do RMFinder se limitavam a avaliar a dispersão dos modelos representativos na curva de risco sem considerar a probabilidade individual de cada modelo representativo. Para calcular $F_{\text{risk}}(S)$, os modelos representativos em S eram ordenados conforme seus valores da variável de produção v e a distância Euclidiana quadrática entre cada modelo representativo é calculada e somada. Entretanto, o RMFinder 2.0 introduz a abordagem de calcular a custo de F_{risk} considerando as probabilidades individuais ao avaliar a sobreposição entre as curvas de risco geradas pelos modelos representativos e pelos modelos originais do problema. A metodologia desta pesquisa considera que os modelos são equiprováveis.
- d) F_{atr} : função objetivo que busca garantir a mesma proporção de níveis de atributo nos modelos representativos. Seu cálculo corresponde a soma quadrática das diferenças entre a quantidade relativa das frequências de cada nível de atributo em uma solução em relação às proporções no conjunto original. Este aspecto não é considerado na metodologia desta pesquisa e, portanto, os conceitos de nível de atributo não são tratados.
- e) F_{freq} : função objetivo que busca garantir a presença dos níveis de atributo nos modelos representativos. Atua basicamente como uma função de penalização que penaliza soluções que não contêm ao menos uma ocorrência de todos os níveis de atributo. F_{freq} retorna o número de níveis de atributos do problema que não estão presentes no conjunto de modelos representativos. Este fator não é considerado na metodologia desta pesquisa.

A configuração utilizada do RMFinder nesta pesquisa gera uma solução de 9 modelos representativos utilizando pesos iguais para cada função objetivo. Foi definido para que os modelos sejam considerados equiprováveis.

3. Resultados

Aplicando o experimento proposto foi obtida uma fronteira de soluções não dominadas composta por 7 soluções. Os MRs que as compõe são expostos na Tabela 1, junto da solução encontrada pelo RMFinder, denominada por S_{RMFinder} .

Na Figura 3 pode-se observar as melhores fronteiras de soluções não dominadas ao longo de quatro gerações distintas, representadas em um gráfico $F_{\text{cluster}} \times F_{\text{distrib}}$. Note que, conforme a

evolução das gerações, as fronteiras se deslocam, indicando a descoberta de soluções progressivamente mais otimizadas. Observe também que algumas soluções estão presentes em fronteiras de diferentes gerações.

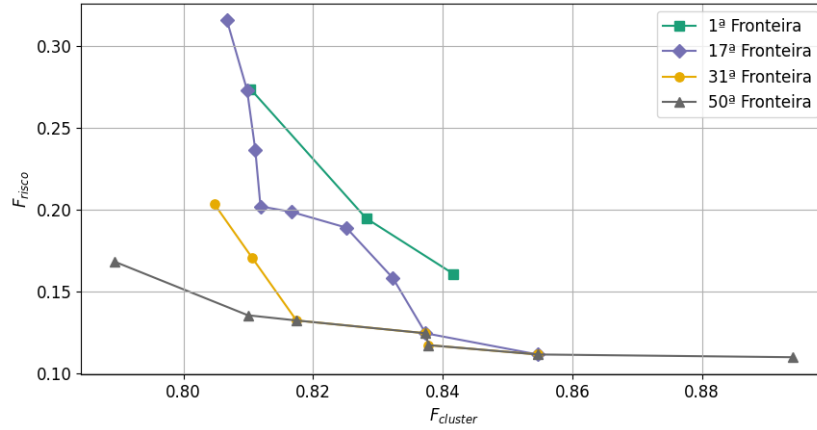


Figura 3: Fronteira de soluções não dominadas ao longo das gerações no plano $F_{cluster} \times F_{distrib}$.

Tabela 1: Solução encontradas pelo RMFinder (primeira linha) e a melhor fronteira encontrada pelo algoritmo proposto (demais linhas).

Solução	Modelos representativos
$S_{RMFinder}$	(19, 44, 77, 106, 142, 144, 151, 164, 177)
S_1	(17, 27, 48, 85, 96, 113, 152, 161, 201)
S_2	(48, 58, 77, 85, 90, 113, 132, 187, 205)
S_3	(23, 25, 33, 45, 94, 117, 120, 175, 187)
S_4	(73, 77, 99, 113, 120, 126, 166, 187, 190)
S_5	(2, 88, 99, 107, 113, 120, 195, 199, 200)
S_6	(67, 84, 86, 93, 122, 140, 157, 166, 212)
S_7	(1, 2, 47, 65, 89, 101, 129, 170, 213)

Na Figura 4, as soluções são representadas pelos custos $F_{cluster} \times F_{distrib}$, cujos valores estão detalhados na Tabela 2. A amplitude desses custos para cada solução não dominada pode ser visualizada na Figura 5. Observa-se uma relação de custo-benefício entre as soluções, onde a redução de um custo está associada ao aumento do outro. Além disso, a amplitude de variação de $F_{cluster}$ foi maior do que a variação de $F_{distrib}$ entre as soluções encontradas.

Com base nas métricas definidas nesta pesquisa, a solução identificada pelo RMFinder, $S_{RMFinder}$, é considerada dominada, uma vez que as soluções S_1 , S_2 e S_3 apresentam custos inferiores para $F_{cluster}$ e $F_{distrib}$.

Com base nas métricas do RMFinder, todas as soluções da fronteira de Pareto apresentam custos maiores que a solução $S_{RMFinder}$, conforme a Tabela 3. Note, porém, que este custo é impulsionado principalmente pelo desempenho em relação às funções objetivo F_{atr} e F_{freq} , que abrange fatores que não foram consideradas no desenvolvimento deste trabalho.

Os resultados demonstram que as soluções encontradas pelo algoritmo proposto por esta

Tabela 2: Custos das funções objetivo F_{cluster} e F_{distrib} .

Solução	F_{cluster}	F_{distrib}
S_{RMFinder}	0.8303	0.1850
S_1	0.7894	0.1681
S_2	0.8099	0.1355
S_3	0.8174	0.1323
S_4	0.8374	0.1245
S_5	0.8378	0.1172
S_6	0.8546	0.1115
S_7	0.8940	0.1099

Tabela 3: Métricas do RMFinder. As funções objetivo F'_{RMFinder} , F'_{cross} , F'_{risk} , F'_{atr} e F'_{freq} possuem seu valor normalizado em relação à solução com menor custo, S_{RMFinder} .

Solução	F_{RMFinder}	F'_{RMFinder}	F'_{cross}	F'_{risk}	F'_{atr}	F'_{freq}
S_{RMFinder}	16.9137	1.00	1.00	1.00	1.00	1.00
S_1	35.2317	2.08	1.17	1.20	1.50	8.00
S_2	25.1527	1.49	1.07	1.13	1.25	4.00
S_3	27.7607	1.64	1.16	1.13	1.28	5.00
S_4	33.2492	1.96	1.16	1.13	1.54	7.00
S_5	27.5900	1.63	1.15	1.08	1.62	4.00
S_6	31.9792	1.89	1.22	1.10	1.63	6.00
S_7	24.2757	1.43	1.34	1.14	1.20	3.00

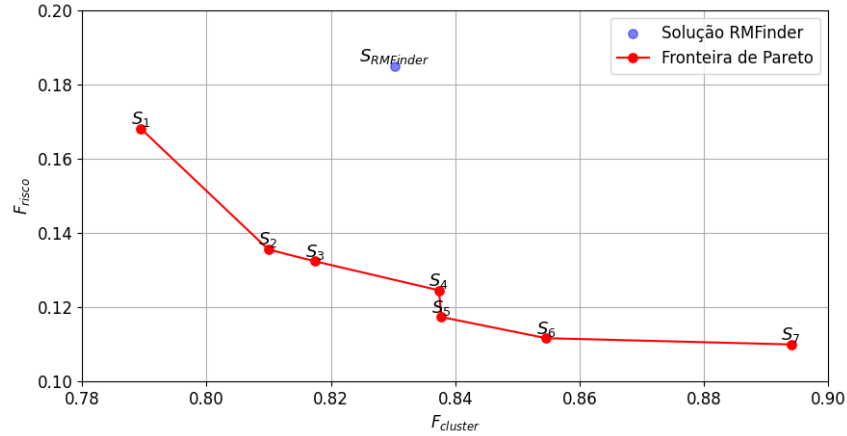


Figura 4: Fronteira de Pareto e solução do RMFinder representados no plano $F_{cluster} \times F_{distrib}$.

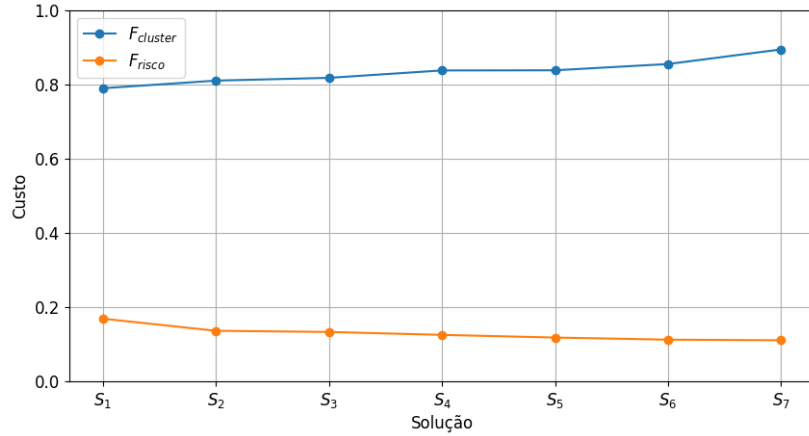


Figura 5: Comparação dos custos $F_{cluster}$ e $F_{distrib}$ para cada solução da fronteira de Pareto.

pesquisa apresentam uma otimização superior para as funções objetivo $F_{cluster}$ e $F_{distrib}$, em comparação com a solução encontrada pelo RMFinder. Por sua vez, a solução $S_{RMFinder}$ mostrou-se mais eficiente na função $F_{RMFinder}$, especialmente em relação às funções objetivo F_{atr} e F_{freq} , que inclui aspectos não considerados neste trabalho. Esse comportamento reflete a diferença de abordagem entre o método deste estudo e o aplicado pelo RMFinder.

4. Conclusão

A definição de uma estratégia de produção adequada na exploração de campos de petróleo exige a análise de diversos modelos geológicos, que, por sua vez, envolvem variáveis incertas. A combinação de todas essas variáveis gera um número exponencial de modelos de reservatório, tornando essencial a seleção de MRs para viabilizar o processamento que determinará a melhor estratégia de produção.

Como um passo de uma metodologia maior, esta pesquisa propôs um método para identificar MRs com base em duas funções objetivo. A primeira avalia a representatividade dos MRs em um espaço $\mathbb{R}^5$, enquanto a segunda mensura a representatividade considerando a distribuição dos MRs nos percentis das variáveis de produção. Foi aplicado o algoritmo genético NSGA-III para minimizar tais funções e encontrar um conjunto de soluções não dominadas.

O algoritmo proposto foi avaliado por meio da aplicação a um benchmark, e seus resultados foram avaliados de maneira cruzada com a ferramenta RMFinder. As soluções geradas por cada metodologia demonstraram superioridade dentro de suas métricas próprias, evidenciando suas diferenças de abordagem. Devido ao problema de seleção de MRs não possuir funções objetivo claramente estabelecidas, os critérios das funções propostas devem ser examinados.

Como trabalho futuro, recomenda-se uma avaliação aprofundada da aptidão das funções objetivo empregadas, bem como o desenvolvimento de novas funções que considerem os aspectos não contemplados nesta pesquisa. Tais aprimoramentos poderão contribuir para uma seleção ainda mais precisa de MRs e para a definição de estratégias de produção mais eficientes.

5. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e com apoio financeiro da do Programa de Pós-Graduação em Tecnologia (PPGT) da Faculdade de Tecnologia da Unicamp.

Referências

- Gaspar, A., Avansi, G., Santos, A., von Hohendorff Filho, J., e Schiozer, D. (2015). Unisim-i-d. benchmark studies for oil field development and production strategy selection. *International Journal of Modeling and Simulation for the Petroleum Industry*, 9:47–55.
- Mahjour, S. K., Correia, M. G., Santos, A. A. d. S. d., e Schiozer, D. J. (2020). Using an Integrated Multidimensional Scaling and Clustering Method to Reduce the Number of Scenarios Based on Flow-Unit Models Under Geological Uncertainties. *Journal of Energy Resources Technology*, 142(6):063005. ISSN 0195-0738. URL <https://doi.org/10.1115/1.4045736>.
- Meira, L. A., Coelho, G. P., da Silva, C. G., Abreu, J. L., Santos, A. A., e Schiozer, D. J. (2020). Improving representativeness in a scenario reduction process to aid decision making in petroleum fields. *Journal of Petroleum Science and Engineering*, 184:106398. ISSN 0920-4105. URL <https://www.sciencedirect.com/science/article/pii/S0920410519308198>.
- Meira, L. A., Coelho, G. P., Santos, A. A. S., e Schiozer, D. J. (2016). Selection of representative models for decision analysis under uncertainty. *Computers & Geosciences*, 88:67–82. ISSN 0098-3004. URL <https://www.sciencedirect.com/science/article/pii/S0098300415300893>.
- Schiozer, D., Ligerio, E., Suslick, S., Costa, A., e Santos, J. (2004). Use of representative models in the integration of risk analysis and production strategy definition. *Journal of Petroleum Science and Engineering*, 44(1):131–141. ISSN 0920-4105. URL <https://www.sciencedirect.com/science/article/pii/S0920410504000403>. Risk Analysis Applied to Petroleum Exploration and Production.
- Schiozer, Denis José, de Souza dos Santos, Antonio Alberto, de Graça Santos, Susana Margarida, e von Hohendorff Filho, João Carlos (2019). Model-based decision analysis applied to petroleum field development and management. *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles*, 74:46. URL <https://doi.org/10.2516/ogst/2019019>.