

LUIZ FELIPE NIEDERMAIER CUSTÓDIO

RECUPERAÇÃO AVANÇADA DE PETRÓLEO: BUSCA DE CAMPOS
ANÁLOGOS E ANÁLISES MUNDIAIS TÉCNICAS E ESTRATÉGICAS
VIA ABORDAGEM ORIENTADA POR CIÊNCIA DE DADOS

São Paulo

2024

LUIZ FELIPE NIEDERMAIER CUSTÓDIO

**Recuperação avançada de petróleo: busca de campos análogos
e análises mundiais técnicas e estratégicas
via abordagem orientada por ciência de dados**

Versão corrigida

Dissertação apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção de título de
Mestre em Ciências.

Área de concentração:
Engenharia Naval e Oceânica

Orientador:
Prof. Dr. Cleyton de Carvalho Carneiro

São Paulo

2024

LUIZ FELIPE NIEDERMAIER CUSTÓDIO

ENHANCED OIL RECOVERY: ANALOGUE OIL FIELDS
AND TECHNICAL AND STRATEGICAL GLOBAL ANALYSIS
THROUGH DATA SCIENCE APPROACH

Master's thesis presented to the Polytechnic School of
the University of São Paulo to obtain the title of
Master of Science degree.

Concentration area:
Naval and Ocean Engineering

Advisor:
Prof. Dr. Cleyton de Carvalho Carneiro

São Paulo

2024

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 16 de junho de 2024

Assinatura do autor: Luiz F N Custódio

Assinatura do orientador: Cleiton de C. Carneiro

Catálogo-na-publicação

Custódio, Luiz Felipe Niedermaier

RECUPERAÇÃO AVANÇADA DE PETRÓLEO: BUSCA DE CAMPOS ANÁLOGOS E ANÁLISES MUNDIAIS TÉCNICAS E ESTRATÉGICAS VIA ABORDAGEM ORIENTADA POR CIÊNCIA DE DADOS / L. F. N. Custódio -- versão corr. -- São Paulo, 2024.

174 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Naval e Oceânica.

1.Recuperação avançada de petróleo 2.Ciência de Dados 3.Análise global de bases de dados 4.Aprendizado de máquina I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Naval e Oceânica II.t.

“Fall in love with some activity and do it! Nobody ever figures out what life is all about, and it doesn't matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough. Work as hard and as much as you want to on the things you like to do the best. Don't think about what you want to be, but what you want to do.” — Richard Feynman

Resumo

Na busca por maximizar a recuperação de reservatórios de petróleo, métodos de Enhanced Oil Recovery (EOR) tornaram-se foco estratégico no setor produtivo petrolífero. No que concerne à análise integrada de tendências globais e fatores determinantes na aplicação destes métodos, no entanto, há uma ampla discussão a ser desenvolvida na literatura científica e tecnológica. Este estudo foi conduzido com uma abordagem global, coletando um vasto conjunto de dados desestruturados de documentos públicos relacionados aos métodos EOR tradicionais e tecnologias emergentes. Assim, três etapas principais foram abordadas: (i) construção de uma ampla base de dados com posição espacial acerca da aplicação de métodos de EOR no mundo; (ii) desenvolvimento de uma análise abrangente e mundial da aplicação de métodos de EOR; (iii) identificação de padrões e principais fatores técnicos, econômicos e regulatórios que influenciam na aplicação dos métodos de EOR no mundo; (iv) desenvolvimento de metodologia para identificação de campos análogos conforme critérios de agências reguladoras. Através da aplicação de técnicas de ciência de dados e da utilização de mapas auto-organizáveis (SOM), foi possível processar, reduzir a dimensionalidade e visualizar a complexidade dos dados. Estas técnicas demonstraram ser particularmente eficazes para dados de alta dimensionalidade e incompletos. O estudo integrou diferentes modalidades de dados, produzindo uma representação profunda do conhecimento inerente aos projetos EOR e suas aplicações em escala global. Os resultados evidenciaram inter-relações sutis entre múltiplas variáveis e os métodos EOR. Revelaram-se tendências, direcionamentos em pesquisa e desenvolvimento, e fatores críticos para decisões informadas. Uma análise criteriosa expôs fatores técnicos, como propriedades dos reservatórios, e não-técnicos, incluindo aspectos locais, comerciais, temporais e regulatórios, que influenciam na seleção e aplicação dos métodos EOR. Foi observada a formação de grupos de similaridade entre países e empresas no que se refere à aplicação de EOR, bem como discussões sobre aspectos regulatórios. Complementarmente, desenvolveu-se uma metodologia eficiente para obtenção de campos análogos, considerando parâmetros e requisitos de agências reguladoras, sendo posteriormente validada com um estudo de caso concreto, e que possibilita decisões mais bem informadas com base em reservatórios com características semelhantes. O estudo também abordou a dinâmica de disponibilidade e custos associados a fluidos de injeção, e seu impacto nas decisões. A exploração de uma base de dados global ofereceu uma compreensão refinada dos fatores que moldam a aplicação global dos métodos EOR, além de introduzir inovações metodológicas e proposições analíticas. O desenvolvimento passa a ser um instrumento útil para profissionais, acadêmicos e tomadores de decisão no domínio da engenharia de reservatórios.

Palavras-chave: Recuperação avançada de petróleo; Ciência de Dados; Análise global de bases de dados; Aprendizado de máquina.

Abstract

In the continuous endeavor to augment oil reservoir recovery, Enhanced Oil Recovery (EOR) methods have risen as a strategic linchpin in the petroleum sector. Despite their undoubted significance, current literature exposes a discernible void in integrating global trends and pivotal determinants driving these methods. Addressing this void, our research embarked on a holistic approach, collating an expansive cache of unstructured data from public documents, spanning both conventional EOR methods and nascent technologies. This endeavor was navigated through stages, commencing with the development of a georeferenced database on global EOR applications, followed by a comprehensive global analysis of EOR methods' deployment, decoding patterns, and the salient technical, economic, and regulatory factors steering their adoption, and culminating in the formulation of a methodology to identify analogous fields aligned with regulatory agency criteria. Leveraging data science methodologies, particularly self-organizing maps (SOM), facilitated the intricate tasks of data processing, dimensionality reduction, and visual interpretation, showcasing their proficiency in handling high-dimensional and sporadic data. The findings illuminated nuanced interdependencies between multifarious variables and EOR methods, unveiling emergent trends, research trajectories, and decision-centric inflection points. A rigorous examination unraveled both technical aspects, like reservoir properties, and non-technical facets, encompassing local, commercial, temporal, and regulatory nuances influencing EOR method selection and deployment. Notably, similarity *clusters* among countries and enterprises regarding EOR application emerged, supplemented by an in-depth discourse on regulatory considerations. An auxiliary achievement was the robust methodology sculpted for analogous field identification, adhering to regulatory guidelines, further authenticated via a tangible case study. This aids in fostering judicious decisions rooted in reservoirs of congruent attributes. Moreover, the dynamics of availability and cost implications tied to injection fluids and their consequent influence on decisions were explored. This research not only furnishes a granulated understanding of the multifaceted ecosystem dictating global EOR methods' adoption but also pioneers methodological breakthroughs and analytical epiphanies, standing as an invaluable beacon for practitioners, scholars, and decision-makers in reservoir engineering.

Keywords: Enhanced Oil Recovery; Data Science; Global databases analysis; Machine Learning.

Lista de figuras

- Figura 1. Diferentes meios possíveis de recuperação de petróleo nas fases de desenvolvimento de um campo de petróleo. Adaptado de Kokal (2012). 1-14
- Figura 2. Número de publicações por ano na base OnePetro relativos a cada categoria de EOR ao longo do tempo e preços do barril de petróleo Brent (nominal e descontada a inflação tendo como base os EUA). Os dados do número de publicações foram obtidos a partir dos resultados de pesquisa no OnePetro e os dos preços de barril de petróleo são provenientes de relatório da U.S. Energy Information Administration (EIA, 2020). O preço descontado a inflação também fora incluído para se obter uma comparação mais justa em cenários econômicos distintos. 1-15
- Figura 3. Preços do barril de petróleo Brent (nominal e descontada a inflação tendo como base os EUA) e número de papers presentes na base OnePetro relativos a cada categoria de EOR ao longo do tempo. Dados do número de papers foram obtidos a partir dos resultados de pesquisa no OnePetro e os dos preços de barril de petróleo são provenientes de relatório da U.S. Energy Information Administration (2019). O preço descontado a inflação também fora incluído para se obter uma comparação mais justa em cenários econômicos distintos. 2-22
- Figura 4. Número de registro de trabalhos relacionados ao tópico "Enhanced Oil Recovery" na base Web of Science, para cada categoria. Verifica-se uma grande interdisciplinariedade em tópicos de pesquisa relacionados ao EOR. 2-29
- Figura 5. Métodos de EOR abordados pelos principais artigos revisados, que possuem resultados na análise de projetos de EOR. São listadas as revistas e ano em que foram publicados, números de citações obtidos no Google Scholar, com ordenação decrescente pelo número médio de citações por ano para cada artigo. Além disso, também é listada uma contabilização com o número de métodos de EOR e de variáveis que cada trabalho aborda, dividindo-se nas suas principais classes, a saber: métodos térmicos, injeção de gases, injeção de produtos químicos, e outros. 3-13
- Figura 6. A Variáveis abordadas pelos principais artigos revisados, que possuem resultados na análise de projetos de EOR. Para cada artigo são listadas as revistas e ano em que foram publicados, números de citações obtidos no Google Scholar, com ordenação decrescente pelo número médio de citações por ano. Além disso, também é listada uma contabilização com o número de métodos de EOR e de variáveis que cada trabalho aborda, bem como na última linha o escopo abordado pelo presente trabalho. 3-14
- Figura 7. Métodos de visualização de dados utilizados pelos principais artigos revisados, que possuem resultados com foco na análise de projetos de EOR. Para cada artigo são listadas as revistas e ano em que foram publicados, números de citações obtidos no Google Scholar, com ordenação decrescente pelo número médio de citações por ano. Além disso, também são listadas as quantidades e abrangência das bases de dados utilizadas pelos artigos e dos países que contemplam, bem como na última linha o escopo abordado pelo presente trabalho. 3-15
- Figura 8. Síntese da rotina metodológica iterativa de mineração de dados, baseada nos modelos KDD e CRISP-DM, utilizada no presente trabalho. 4-13
- Figura 9. Síntese da aplicação do algoritmo SOM na base de dados estruturada obtida no presente trabalho, constituído das etapas: (a) base de dados processada, com as variáveis selecionadas e modeladas, é introduzida no algoritmo SOM; (b) o algoritmo SOM é inicializado, ocorrendo o processo de aprendizado e ajuste dos BMUs no espaço n-dimensional; (c e d) a rede neural tem os seus pesos projetados no hipertoróide e a U-Matrix (e) as *component plots* (f) são obtidas a partir da planificação dos respectivos hipertoróides. A U-Matrix representa as dissimilaridades no ambiente SOM dos BMUs e as *component plots* a influência de cada variável para cada conjunto de amostras representadas por um BMU no espaço SOM. 4-17

- Figura 10. Rotina metodológica desenvolvida para a obtenção e análise de campos análogos. 4-19
- Figura 11. Representação das medidas de similaridade distância euclidiana e ângulo do cosseno para dois pontos A e B. 4-23
- Figura 12. Ilustração de agrupamento realizado por clusterização hierárquica e sua representação na forma de dendrogramas a partir de amostras dispostas no espaço bidimensional. 4-24
- Figura 13. Constituição da base de dados em relação às classes de métodos de EOR presentes. 5-26
- Figura 14. Localização dos campos de petróleo que tiveram aplicação de EOR presentes na base de dados aqui apresentada. 5-27
- Figura 15. Barras de frequência de projetos para cada categoria de EOR ao longo dos anos sobreposta pela curva relativa ao preço do barril de petróleo Brent, nominal e descontada a inflação, tendo como base índice econômico dos EUA. 5-28
- Figura 16. Categorical scatterplot com a distribuição das subclasses de métodos de EOR vs os países em que foram aplicados. O tamanho dos pontos é proporcional à frequência acumulada de projetos do método para com o respectivo país. 68
- Figura 17. *Treemap* representando a distribuição dos projetos de EOR, agrupados conforme classe de EOR, o país de localização e classificação da empresa operadora do campo, conforme ela seja IOC (International Oil Company), Big (receita anual entre 5 e 20 bilhões de dólares), Medium (Medium, receita anual entre 500 milhões e 5 bilhões de dólares) e Small (receita anual abaixo de 500 milhões de dólares). A cor de cada retângulo refere-se à data médio de início dos projetos, indicada pela escala ao lado. 5-30
- Figura 18. *Scatter plot* das variáveis profundidades dos reservatórios e datas de início dos projetos, com categorização dos dados quanto à classe do método de EOR utilizada. 5-31
- Figura 19. Joinplot com a plotagem cruzada das variáveis mobilidade estimada do óleo e tamanho estimado da garganta de poro. As retas r) e s) indicam tendências de correlação distintas para subgrupos de dados. 5-33
- Figura 20. *Boxplots* para os principais tipos de métodos de EOR para diferentes variáveis de natureza técnica e petrofísica. 5-34
- Figura 21. *Treemap* representando a distribuição dos projetos de EOR, agrupados conforme classe (injeção de gás, térmico, químico ou outros), o país de localização e classificação da empresa operadora do campo, conforme ela seja IOC (International Oil Company), B (Big, com receita anual entre 5 e 20 bilhões de dólares), M (Medium, com receita anual entre 500 milhões e 5 bilhões de dólares) e S (Small, com receita anual abaixo de 500 milhões de dólares). Além disso, a cor de cada retângulo refere-se à data médio de início dos projetos, indicada pela escala ao lado. 5-35
- Figura 22. *Treemap* representando a distribuição dos projetos de EOR, agrupados conforme hierarquia de classe (injeção de gás, térmico, químico ou outros), do país de localização e da empresa operadora do campo. Além disso, a cor de cada retângulo refere-se à data médio de início dos projetos, indicada pela escala ao lado. 5-36
- Figura 23. *Treemap* exibindo uma representação hierarquizada dos dados relativos às empresas operadores dos projetos de EOR, agrupados pelas classes (métodos de injeção de gás, térmicos, químicos e outros) e respectivos países em que estão localizados tais campos. Cada retângulo, representando uma dada empresa, tem seu tamanho proporcional à quantidade de projetos que tal empresa possui. Além disso, as cores dos retângulos, conforme a escala, remetem à profundidade média dos reservatórios em que se teve aplicação de EOR para uma dada classificação. 5-36
- Figura 24. *Boxplots* para as variáveis a) área em acres do projeto, b) e c) número de poços produtores e injetores utilizados, e d) e e) para as respectivas densidades de poços nos

projetos de EOR. Os dados são segmentados, em um primeiro nível, em relação às classes de métodos de EOR utilizados. Em um segundo nível, são categorizados quanto ao porte e atuação das companhias operadoras, podendo ser pequenas empresas (Small), médias empresas (Medium), grandes empresas (Big), companhias nacionais e/ou estatais – National Oil Companies (NOCs), e companhias privadas com atuação internacional – International Oil Companies (IOCs). 5-38

Figura 25. Matriz-U resultante do treinamento realizado com os dados de projetos de EOR, com quatro níveis de destaques: a) Matriz-U com representação da rede hexagonal de neurônios, onde cada contorno branco externo de hexágono representa um neurônio da rede neural artificial, e em que os hexágonos adjacentes à borda indicam a medida de dissimilaridade. As cores mais quentes indicam altas dissimilaridades do neurônio em questão com os seus adjacentes, e as cores mais frias maiores similaridades; b) Matriz-U com hits representados na cor branca. O tamanho do hexágono é proporcional à quantidade de reservatórios que o respectivo BMU está representando; c) linhas interpretadas relacionadas às regiões de neurônios de maior influência exercida pelos dados do respectivo método de EOR; e d) sobreposição das regiões das linhas interpretadas em c) na imagem da Matriz-U de a). 5-39

Figura 26. Matriz-U resultante do treinamento realizado com os dados de projetos de EOR. Exemplos de projetos representativos dos respectivos BMUs possuem seus nomes apresentados. 5-40

Figura 27. a) *Component plots* referente ao método de injeção de vapor; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor. 5-42

Figura 28. a) *Component plot* referente ao método de injeção de vapor; b) a f) *component plots* de variáveis relacionadas às litologias dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor. 5-43

Figura 29. a) *Component plot* referente ao método de injeção de vapor; b) a d) *component plots* referentes à escala do projeto (se piloto, escala de reservatório ou escala de campo); e) a Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor referente a injeção de vapor. 5-44

Figura 30. Boxplot referente às variáveis data de início do projeto de EOR e preço do barril do petróleo descontado a inflação para a data de início. Segmentação dos dados quanto ao tipo de método térmico. 5-46

Figura 31. a) *Component plot* referente ao método de injeção de vapor; b) a d) *component plots* referentes à escala do projeto (se piloto, escala de reservatório ou escala de campo); e) a Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor referente a injeção de vapor. 5-46

Figura 32. Boxplot referente às variáveis a) área em acres, b) número de poços produtores, c) número de poços injetores, e densidade de poços d) produtores e e) injetores. Segmentação dos dados, num primeiro nível, quanto ao método térmico utilizado para cada boxplot, e quanto ao porte de atuação das empresas operadoras como diferentes cores para os pontos. 5-47

Figura 33. Boxplot referente às variáveis a) profundidade e b) °API. Segmentação dos dados, num primeiro nível, quanto ao método térmico utilizado para cada boxplot, e quanto ao porte de atuação das empresas operadoras como diferentes cores para os pontos. 5-48

Figura 34. a) *Component plot* referente ao método combustão *in-situ*; b) a f) *component plots* referentes à escala do projeto (se piloto, escala de reservatório ou escala de campo). A linha em destaque delimita os neurônios ativados por reservatórios associados ao método de combustão *in-situ*. 5-49

Figura 35. a) *Component plot* referente ao método de combustão *in-situ*; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de combustão *in-situ*. 5-50

Figura 36. a) *Component plot* referente ao método de injeção de gás hidrocarboneto; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás hidrocarboneto. 5-51

Figura 37. a) *Component plot* referente ao método de injeção de gás hidrocarboneto; b) e c) *component plots* de variáveis relacionadas às litologias dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás hidrocarboneto. 5-52

Figura 38. *Boxplots* referente às variáveis data de início do projeto de EOR e preço do barril do petróleo descontado a inflação para à data de início. Segmentação dos dados quanto às subclasses de métodos de EOR utilizadas. 5-53

Figura 39. a) *Component plot* referente ao método de gás hidrocarboneto; b) o preço do barril ajustado pela inflação no momento de início dos projetos; c) se o projeto é *offshore*; e) a f) *component plots* referentes à escala do projeto (se implementado na forma de projeto piloto, em escala de reservatório ou em escala de campo). 5-54

Figura 40. a) *Component plot* referente ao método de injeção de gás hidrocarboneto; b) e c) *component plots* referentes ao segmento de atuação (se integrada na cadeia de valor ou independente e focada na exploração e produção – *Upstream*); d) e e) tamanho da companhia operadora, se International Oil Company (IOC) ou empresa de grande porte, mas com atuação regional; f) Matriz-U. 5-56

Figura 41. a) *Component plot* referente ao método de injeção de gás CO₂; b) e c) *component plots* relacionadas a se o processo de injeção é miscível ou imiscível; d) a f) *component plots* de variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás CO₂ 5-58

Figura 42. a) *Component plot* referente ao método de injeção de gás CO₂; b) a d) *component plots* de variáveis relacionadas a variáveis básicas dos reservatórios; e) e f) *component plots* de variáveis relacionadas às litologias dominantes dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção gás CO₂. 5-59

Figura 43. a) *Component plot* referente ao método de injeção de gás CO₂; b) *component plot* referente ao segmento de atuação (se independente e focada na exploração e produção – *Upstream*); c) a f) tamanho da companhia operadora (se empresa de grande porte, mas com atuação regional; se empresas de pequeno e médio porte. e se International Oil Company (IOC)). 5-61

Figura 44. a) *Component plot* referente ao método de injeção de gás nitrogênio; b) e c) em relação a se o processo de injeção do gás é miscível ou imiscível; d) a f) *component plots* de variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás nitrogênio. 5-63

Figura 45. a) *Component plot* referente ao tipo de gás injetado nitrogênio; b) e c) *component plots* referentes às litologias dominantes dos reservatórios; d) Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás nitrogênio. 5-64

Figura 46. a) *Component plot* referente ao método de gás nitrogênio; b) e c) número de poços produtores e injetores; d) se o projeto é *offshore*; e) ano de início do projeto de EOR; f) o preço do barril ajustado pela inflação no momento de início dos projetos. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás nitrogênio. 5-65

Figura 47. a) *Component plot* referente ao método de injeção de polímero; b) a f) *component plots* de variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero. 5-67

Figura 48. a) *Component plot* referente a injeção de polímero; b) a d) *component plots* referentes às litologias dominantes dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero. 5-68

Figura 49. *Boxplots* para as variáveis a) TDS e b) transmissibilidade de acordo categorizados quanto a métodos de EOR. 5-69

Figura 50. a) *component plot* referente ao tipo de produto químico injetado polímero; b) se o projeto é *offshore*; c) a e) à profundidade e temperatura dos respectivos reservatórios e suas datas inícios de execução; f) preços do barril ajustados pela inflação no início do projeto. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero. 5-70

Figura 51. Para os dados de projetos que tiveram injeção de métodos químicos, scatterplots da variável profundidade com a) preço ajustado do barril preço ajustado do barril e b) data de início do projeto; da variável temperatura com c) preço ajustado do barril preço ajustado do barril e d) data de início do projeto. Para os dados de projetos que tiveram projetos com métodos de injeção de gás, scatterplots da variável profundidade com e) preço ajustado do barril preço ajustado do barril e f) data de início do projeto. 5-72

Figura 52. a) *component plot* referente ao tipo de químico injetado (polímero); b) e c) *component plots* referentes à escala do projeto (se implementado na forma de projeto piloto, em escala de reservatório ou em escala de campo); d) Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero. 5-74

Figura 53. a) *Component plot* referente ao método de injeção de polímero; b) *component plot* referente ao segmento de atuação (se integrada na cadeia de valor); d) e e) tamanho da companhia operadora, se International Oil Company (IOC) ou National Oil Company (NOC). A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero. 5-75

Figura 54. a) *component plot* referente ao método de injeção de surfactante e polímero; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de surfactante e polímero. 5-77

Figura 55. a) *component plots* do tipo de químico injetado (surfactante e polímero), b) e c) data de início e preço ajustado do barril no ano de início dos projetos; d) Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de surfactant e polímero. 5-78

Figura 56. *Component plots* referente a a) injeção microbial e b) a f) variáveis de propriedades básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção microbial. 5-80

Figura 57. *Component plots* referente a a) injeção microbial e b) a d) variáveis referentes às litologias dominantes dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção microbial. 5-81

Figura 58. *Component plots* referente a a) injeção microbial; b) implementação de projetos na China, c) ano de início dos projetos; d) se a empresa operadora de cada reservatório é classificada como National Oil Company. 5-83

Figura 59. *Component plots* referentes a a) método de injeção de água de baixa salinidade (LSWF); b) a f) propriedades básicas dos reservatórios. 5-84

Figura 60. *Component plots* referentes a a) método de injeção de água de baixa salinidade (LSWF); b) a d) litologias dominantes dos reservatórios. 5-85

Figura 61. *Component plots* referentes a a) método de injeção de água de baixa salinidade (LSWF); b) data de início do projeto; c) preço do barril no início do projeto ajustado pela inflação; d) Matriz-U. 5-86

Figura 62. Clusterização dos países identificados com projetos de EOR e representação na forma de dendrogramas, com a identificação de cinco grupos principais de países. 5-88

- Figura 63. Clusterização das empresas identificadas com projetos de EOR e representação na forma de dendrogramas, com a identificação de três grupos principais de empresas 5-90
- Figura 64. Visualização da relação entre as *components* 1 e 2 com as variáveis do conjunto de dados de campos análogos. 5-91
- Figura 65. Histogramas comparando os dados do campo de interesse e do conjunto de candidatos a análogos com a base de dados geral para os valores médios das variáveis numéricas a) porosidade, b) permeabilidade, c) profundidade, d) °API, e) viscosidade e f) temperatura de cada reservatório. 5-93
- Figura 66. Scatterplots para as variáveis numéricas permeabilidade, porosidade, profundidade, °API e viscosidade do óleo. Os pontos são categorizados quanto a se o dado é referente ao reservatório de interesse (em vermelho), se faz parte do conjunto de candidatos avançados (em verde), se faz parte do conjunto inicial de candidatos (em laranja), ou se da base de dados inicial. 5-94
- Figura 67. Distribuição dos dados, no formato de polígonos, de acordo com as variáveis numéricas utilizadas. Os dados são categorizados quanto a se é do reservatório target (em amarelo), candidato a análogo (em verde) ou da base de dados geral (em cinza). As variáveis permeabilidade e viscosidade possuem escalas logarítmicas. 5-95
- Figura 68. Matriz de clusterização hierárquica obtida utilizando agrupamento aglomerativo para o conjunto de reservatórios candidatos a análogos ao campo de Serra. Cada linha se refere a um campo da base de dados de análogos candidatos, e as colunas a cada variável considerada na busca de análogos. 5-97
- Figura 69. Dendrograma circular obtido a partir da clusterização hierárquica em relação às variáveis numéricas dos campos candidatos a análogos. 5-98
- Figura 70. Visualização bidimensional das duas primeiras *components* principais obtidas pela aplicação da técnica PCA no conjunto de dados em questão. Categorização dos dados quanto a se consistem dos reservatório target, dos campos análogos obtidos ou da base de dados geral. 5-99
- Figura 71. Grupos de reservatórios análogos internacionais semelhantes ao campo de Serra. 5-100
- Figura 72. Quadro comparativo de características geológicas do campo de Serra com o campo de West Hastings, da bacia da Costa do Golfo. 5-101
- Figura 73. Quadro comparativo de características geológicas do campo de Serra com o campo de Hawkings, da bacia do East Texas 5-102
- Figura 74. Quadro comparativo de características geológicas do campo de Serra com o campo de Loudon, da bacia do Illinois. 5-103

Lista de tabelas

Tabela 1. Pseudocódigo do algoritmo <i>Self-Organizing Maps</i> , explicitando os principais parâmetros de entrada, processo de inicialização, treinamento e aprendizagem. (Autoria própria)	3-39
Tabela 2 - Hiperparâmetros de inicialização do algoritmo SOM.	4-25

Lista de abreviaturas e siglas

BMU	Best Matching Unit
CCS	Carbon, Capture and Storage
CRISP-DM	Cross-Industry Standard Process for Data Mining
EOR	Enhanced Oil Recovery
HC	Hydrocarbon
HWF	Hot Water Flooding
IOC	International Oil Company
ISC	In-Situ Combustion
KDD	Knowledge Discovery in Databases
LSWF	Low Salinity Waterflooding
NOC	National Oil Company
SAGD	Steam-Assisted Gravity Drainage
SOMs	Self-Organizing Maps
SP	Surfactante and Polymer
WAG	Water Alternating Gas

Sumário

1. INTRODUÇÃO	1-13
1.1. Objetivos	1-17
1.1.1. <i>Objetivos específicos</i>	1-17
2. RELEVÂNCIA DA PESQUISA	2-18
2.1. Quanto à competitividade e demanda energética	2-18
2.2. Quanto à difusão do uso do EOR	2-19
2.3. Quanto ao uso de Bases de Dados Integradas em algoritmos de Inteligência Artificial	2-26
2.4. Quanto ao potencial de aplicação de EOR no Brasil	2-27
2.5. Quanto à interdisciplinaridade na aplicação de EOR	2-28
2.6. Quanto à importância de uma análise mundial da aplicação dos métodos de EOR	2-29
2.7. Quanto à importância de obtenção de campos análogos	2-30
2.8. Quanto à importância do <i>screening</i> de EOR	2-31
3. FUNDAMENTOS TEÓRICOS	3-33
3.1. Mineração de dados	3-33
3.2. Técnicas de visualização de bases de dados de alta dimensionalidade	3-34
3.3. Self-Organizing Maps	3-35
3.4. Trabalhos prévios envolvendo análise de projetos de EOR	3-40
4. METODOLOGIA	4-13
4.1. Estruturação da Base de Dados	4-14
4.2. Pré-processamento	4-18
4.2.1. <i>Limpeza</i>	4-19
4.2.2. <i>Gerenciamento de variáveis</i>	4-21
4.3. Processamento e metodologia de análise	4-22
4.3.1. <i>Análise Exploratória e visualização dos dados</i>	4-22
4.3.2. <i>Análises Multidimensionais por Inteligência Artificial – Self-Organizing Maps</i>	4-23
4.4. Interpretações e validações de padrões observados	4-25
4.5. Obtenção de campos análogos	4-26
4.5.1. <i>Rodadas avançadas de obtenção de campos análogos</i>	4-28
5. RESULTADOS	5-30

5.1. Visão geral da base de dados	5-30
5.1.1. <i>Localização e histórico dos projetos de EOR</i>	5-30
5.1.2. <i>Variáveis técnicas</i>	5-34
5.1.3. <i>Variáveis de projeto</i>	5-42
5.2. Análises por SOM dos métodos de EOR	5-49
5.2.1. <i>Métodos térmicos</i>	5-50
5.2.2. <i>Injeção de gases</i>	5-65
5.2.3. <i>Injeção de produtos químicos</i>	5-92
5.2.4. <i>Outros métodos</i>	5-105
6. DISCUSSÕES	5-115
6.1. Quanto à metodologia empregada	5-116
6.2. Quanto aos resultados e padrões identificados	5-117
6.2.1. <i>Influência de variáveis técnicas nos métodos de EOR</i>	5-117
6.2.2. <i>Influência de aspectos locais na escolha de métodos de EOR</i>	5-121
6.2.3. <i>Influência de variáveis comerciais nos métodos de EOR</i>	5-122
6.2.4. <i>Similaridades globais entre os métodos de EOR</i>	5-125
7. CONSIDERAÇÕES FINAIS	5-125
8. PRÓXIMAS ETAPAS	8-127
8.1. Cronograma da pesquisa	8-127
9. REFERÊNCIAS BIBLIOGRÁFICAS	9-129
10. ANEXOS	10-135
10.1. Anexo A – Publicação sobre obtenção de campos análogos	10-135

1. INTRODUÇÃO

Os métodos de recuperação avançada de petróleo - do inglês, Enhanced Oil Recovery (EOR) - consistem nos principais meios para aumentar a eficiência da mobilização do óleo remanescente, prolongando a vida útil dos campos de petróleo e promovendo um incremento substancial na produção. Os métodos de EOR, também chamados de recuperação terciária, consistem na extração de petróleo de um campo a partir da injeção de fluidos. Tais fluidos injetados, além de fornecer energia extra ao reservatório, podem também modificar as propriedades do petróleo na interação rocha-fluido, a fim de extrairlo de forma mais eficiente (ABIDIN; PUSPASARI; NUGROHO, 2012; WEVER; PICCHIONI; BROEKHUIS, 2011).

Existem quatro principais classes de métodos de EOR: (1) métodos térmicos; (2) injeção de gases; (3) injeção de produtos químicos; e (4) outros (ALVARADO; MANRIQUE, 2010a). Tais classes apresentam subdivisões em uma série de diferentes métodos de EOR associados, conforme exemplifica a Figura 1. Cada um dos métodos de EOR tende a ser aplicado a determinados tipos de reservatório, não havendo um método que seja globalmente o melhor. O sucesso da implementação desses métodos é dependente tanto das diversas propriedades de rochas e fluidos do reservatório, como também do modelo estrutural, de fatores geográficos, logísticos, econômicos, dentre outros. Além disso, pode haver a combinação de um ou mais métodos de EOR na aplicação em um campo de petróleo. Tal fato gera uma grande gama de possibilidades de combinação de métodos, o que pode tornar a avaliação e seleção dos métodos de EOR uma tarefa ainda mais complexa.



Figura 1. Diferentes meios possíveis de recuperação de petróleo nas fases de desenvolvimento de um campo de petróleo. Adaptado de Kokal (2012).

Por outro lado, o volume e a variedade de dados gerados e disponíveis de forma pública estão aumentando a uma taxa sem precedentes. É crescente também a quantidade de artigos científicos e relatórios confiáveis elaborados por consultorias e órgãos administrativos de vários países voltados para segurança energética abordando a aplicação de métodos de EOR.

Diversos documentos públicos costumam conter dados, ainda que esparsos e de maneira isolada e não estruturada, acerca da aplicação de métodos de EOR em campos de petróleo no mundo inteiro. Tais fontes de dados, além de reunirem informações sobre campos e métodos mais tradicionais de EOR, como injeção de vapor e injeção de hidrocarboneto em reservatórios convencionais, por vezes também disponibilizam informações sobre novas tecnologias, a exemplo do Low Salinity Waterflooding (LSWF). Ainda que em diferentes formatos, bastante fragmentados e dispersos, essas diferentes fontes de dados, com o devido tratamento e processamento, possuem amplo potencial informativo. Estas informações reunidas possibilitam a construção de uma base de dados mundial, abrangente e confiável, com relevantes informações sobre EOR, capaz de promover estudos dirigidos utilizando técnicas avançadas de análises.

A possibilidade de se investigar um amplo conjunto de experiências prévias de aplicação de EOR pode trazer respostas para questões relevantes relacionadas ao desenvolvimento de campos maduros. Há vários trabalhos publicados nos últimos anos nesse sentido, seja abordando EOR de forma mais genérica (BABADAGLI, 2007, 2020; MANRIQUE, 2009; MUGGERIDGE et al., 2014; THOMAS, 2008) ou alguma subclasse de método de recuperação em específico ou com ênfase em alguma determinada região do mundo (BELHAJ et al., 2020; FAKHER; IMQAM, 2020; GBADAMOSI et al., 2018, 2019; OLAJIRE, 2014; ZHANG; WEI; BAI, 2018). Tais trabalhos anteriores, a partir de experiências prévias de EOR trazem detalhes e lições relevantes para a indústria acerca da implementação de EOR. No entanto, a análise e a interpretação de tais dados de maneira holística permanece um desafio significativo. A visualização e extração sintetizada de informações relevantes e significativas a partir de bases de dados com número de dimensões elevado é por vezes notoriamente complexa e complicada (BERTINI; TATU; KEIM, 2011). Além disso, a abrangência dos dados utilizados por esses trabalhos como base por vezes é limitada, seja em relação aos métodos de EOR e às regiões do mundo que contemplam, ou mesmo em relação às variáveis elencadas para análise ou à diversidade de fontes utilizadas. Verifica-se ainda que há uma lacuna na literatura científica e tecnológica quanto a métodos de visualizações multidimensionais dos dados, à integração de diferentes formas e fontes de dados no âmbito de recuperação avançada de petróleo e à utilização dessas informações para identificação de cenários e campos análogos de aplicação e de suporte a tomadas de decisões em projetos de EOR.

1.1. Objetivos

Como objetivo principal, esta pesquisa visa analisar, selecionar e prever parâmetros e métodos de recuperação mais adequados para cada campo de produção, tomando como base projetos reais da aplicação de EOR em escala mundial, integrados a partir de uma base de dados única, processada por algoritmos de Inteligência Artificial. Como abordagens principais pretende-se (i) analisar a dinâmica das variáveis que influenciam nos métodos de EOR e (ii)

suportar tomadas de decisões no contexto de aplicação de EOR por meio de abordagens semi-automatizadas orientadas por dados.

1.1.1. Objetivos específicos

No contexto da análise das variáveis que influenciam na aplicação dos métodos de EOR, a pesquisa se propõe a:

- Construir uma ampla base de dados georreferenciada acerca da aplicação de métodos de EOR no mundo;
- Desenvolver uma análise abrangente e mundial da aplicação de métodos de EOR;
- Analisar a interrelação entre variáveis e identificação de padrões;
- Identificar os principais fatores técnicos ou não que influenciam na aplicação dos métodos de EOR.

Em relação ao suporte a tomadas de decisões no contexto de aplicação de EOR, pretende-se:

- Desenvolver uma sistemática para obtenção de campos análogos a partir de ampla base de dados mundial de experiências de produção;

2. RELEVÂNCIA DA PESQUISA

2.1. Quanto à competitividade e demanda energética

Dentre as fontes de fornecimento de energia primária no mundo, petróleo e gás natural representam 35% e 28%, respectivamente, da matriz energética mundial atual (IEA, 2022). Atualmente, várias tecnologias baseadas em energia renovável estão sendo desenvolvidas e propostas para reduzir nossa dependência de combustíveis fósseis, como energia solar e eólica, biocombustíveis (bioetanol, biodiesel, hidrogênio, etc.), células de combustível microbiana e fontes de energia alternativas. No entanto, a maioria das tecnologias ainda são prematuras ou carecem de competitividade econômica e sua adoção tem se baseado essencialmente a partir de estímulos governamentais (CHEN et al., 2018; CHOI, 2015; CIPRIANI et al., 2014; DASAN et al., 2019; GEBREMARIAM; MARCHETTI, 2018; GOH et al., 2019; LYND, 2017; MAHLIA et al., 2020; MOHD AZHAR et al., 2017; ZABED et al., 2017). Assim, o petróleo ainda é para muitos países, em especial para as nações em desenvolvimento, a forma de energia mais acessível e econômica. Somado a isso, produtos derivados do petróleo suportam também toda uma cadeia de fornecimento de combustíveis e de produtos químicos diversos que constituem uma das bases da sociedade moderna.

Os efeitos de quaisquer mudanças drásticas nos preços do petróleo bruto podem ser vistos globalmente e podem refletir nas relações, mercados e políticas internacionais. Nos últimos anos, especialmente com a crise causada pelo COVID-19 temos visto os preços do petróleo oscilantes devido a desequilíbrios na 'oferta e demanda do mercado'. Tal fato levou a uma redução no preço do barril de petróleo, que se espera permanecer na faixa de 40 a 70 USD por um período prolongado (MCKINSEY, 2020). O cenário atual tão volátil para os preços do petróleo, por um lado, torna economicamente mais atrativo o uso do petróleo como fonte de energia básica por diversos setores em detrimento de alternativas renováveis que em geral ainda possuem um alto custo (MCCOLLUM et al., 2016). Por outro lado, é um caminho bastante árduo para as indústrias de petróleo e para os países que são altamente dependentes da

receita proveniente do petróleo. Nesse cenário, hoje, mais do que nunca, e no contexto da Indústria 4.0, a resiliente indústria do petróleo busca formas de continuar se reinventando frente a novos desafios. Somado a isso, o petróleo é um recurso finito, com reservas declinantes; a capacidade de manter sua produção em diversas regiões do globo tem sido uma preocupação recorrente por mais de 50 anos (MILLER; SORRELL, 2014) e a descoberta de novas reservas é um processo cada vez mais complexo (KANG; LIM; HUH, 2016; NINGNING; WANG; JIANJUN, 2018). Dentre as formas de se aumentar a competitividade da indústria e, principalmente, suprir o abastecimento de petróleo para as próximas décadas, está o emprego de métodos avançados de recuperação com o intuito de maximizar a eficiência da produção de óleo dos reservatórios (JIN, 2017; KANG; LIM; HUH, 2016; REKSIDLER et al., 2015).

2.2. Quanto à difusão do uso do EOR

A busca por alternativas que permitam uma melhor eficiência no processo de produção do petróleo cresce à medida em que aumenta a sua respectiva demanda. Da mesma maneira, buscam-se alternativas para aumento da recuperação de petróleo em campos maduros, à medida em que diminuem as reservas de petróleo *onshore* e de margem continental. Ao longo do ciclo de vida de um campo, os reservatórios perdem a capacidade de produzir óleo naturalmente, por meio de sua energia primária, o que acarreta em considerável quantidade de óleo não recuperado. Outrossim, parte do petróleo produzido no mundo vem de campos de petróleo maduros, que estão na segunda metade de seus ciclos de produção e, em muitos casos, a descoberta e exploração de novas reservas pode ser dispendiosa e demorada. Conforme afirmam (ALVARADO; MANRIQUE, 2010a), dos cerca de 3 trilhões de barris de petróleo conhecidos em reservatórios convencionais, cerca de menos de um terço foi produzido e consumido no mundo desde o início da exploração e produção de petróleo. Espera-se que um terço adicional do petróleo existente seja produzido por técnicas além das atividades tradicionais de petróleo e gás. Dentre essas técnicas, despontam os métodos de EOR que, apesar da demanda de investimentos, são comercialmente viáveis. A figura 2 ilustra o funcionamento

geral de aplicação de métodos de EOR, nesse CO₂ como o fluido injetante, com a estimulação do reservatório e aumento da produção de óleo

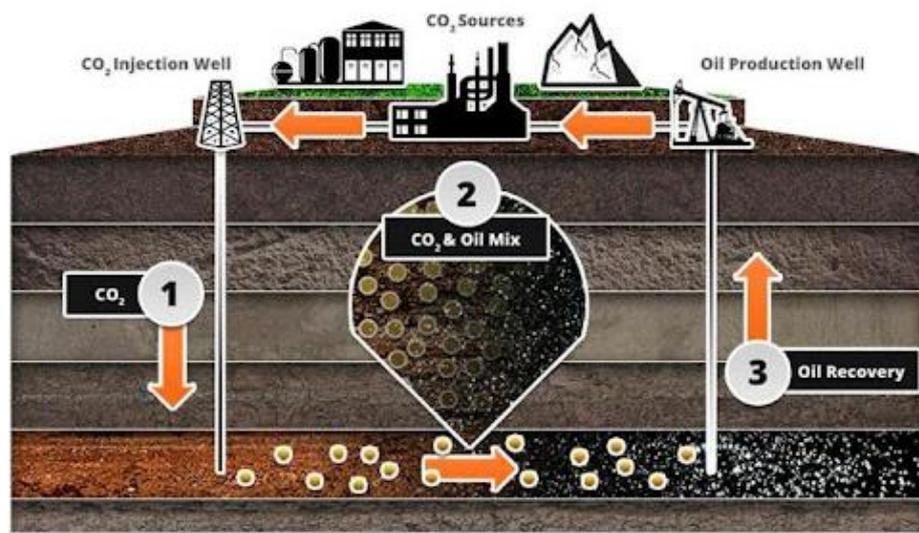


Figura 2. Ilustração da aplicação de métodos de EOR, em particular da injeção de CO₂ com as seguintes principais etapas: (1) transporte de uma fonte externa de CO₂ via gasodutos de indústrias emissoras para os poços injetores, podendo ser injetado com água ou outros aditivos de forma conjunta; (2) etapa em que o fluido injetado entra em contato com o óleo, promovendo mecanismos de deslocamento, miscibilização, redução da tensão interfacial e/ou volatilização do óleo; (2) óleo sendo produzido e então enviado para o mercado, com o CO₂ e água produzidos podendo ser tratados e reinjetados no reservatório. Figura extraída e adaptada de Core Energy LLC (2022).

De acordo com KOKAL (2010), a dinâmica que governa a eficiência dos métodos de EOR é, todavia, complexa e demonstra ser desafiadora, tanto tecnicamente como economicamente, exigindo a avaliação de uma ampla gama de parâmetros. Devido a fatores como a alta sensibilidade técnica, o elevado custo envolvido, a dependência de fatores externos e a complexidade de análise, têm-se buscado novos meios de realizar a seleção dos métodos de EOR mais adequados para cada reservatório. É frequente nas empresas, no entanto, um dilema entre concentrar investimentos na exploração, com o propósito de encontrar novos campos de petróleo, ou aplicar métodos de EOR a campos já maduros. Somado a isso, muitas vezes as incertezas geológicas e econômicas associadas à exploração de novos campos costumam ser muito elevadas. Em

campos maduros, o conhecimento acerca do comportamento dos respectivos reservatórios tende a possibilitar a aplicação de EOR com menores incertezas, diminuindo os riscos financeiros e ambientais associados. Contudo, muitas empresas tendem a não aplicar métodos avançados de recuperação em reservatórios que, em princípio, possuiriam grandes potenciais de ganhos.

Historicamente, a implementação de métodos EOR era por parte das operadoras uma tomada de decisão tardia, considerada apenas quando a produção proveniente de métodos de recuperação primários e secundários já se encontravam exauridos. Alguns fatores podem contribuir para que as companhias não utilizem os métodos de EOR como prática comum, a saber: (i) complexidade dos fenômenos envolvidos; (ii) ausência de conhecimento técnico especializado; (iii) demanda por pesquisa e estudos mais detalhados acerca da aplicação das tecnologias de EOR no portfólio de reservatórios da empresa; (iv) possibilidade de explorar novas reservas com *paybacks* menores, e (v) a falta de recomendações específicas de possíveis métodos de EOR candidatos à aplicação, dentre outros fatores (MANRIQUE, 2014). Hoje em dia, a maioria dos campos de produção no mundo apresentam curvas de produções em declínio. Nesse cenário, as longas investigações tradicionais de EOR não se encaixam no fluxo de trabalho da economia moderna em ampla variedade de campos. O fato é que, para atingir o fator de recuperação máximo, torna-se necessário o planejamento da aplicação de EOR nos estágios iniciais de desenvolvimento de um campo, bem como monitorar os eventos à medida em que estes acontecem, para maior controle e otimização do resultado (SCHLUMBERGER, 2019).

Segundo MANRIQUE (2014) e MUGGERIDGE et al. (2018), a falta de equipes de engenheiros e geocientistas pode estar associada ao colapso do preço do petróleo nos anos 80 e 90 e à eliminação progressiva de centros de pesquisa e desenvolvimento nas principais empresas de petróleos. Conforme pode-se observar na Figura 3, que relaciona o número de *papers* presentes na base *OnePetro* relativos a cada categoria de EOR com o preço de petróleo, de maneira geral, verifica-se, especialmente até os anos 2010, uma tendência de aumentos/reduções no preço de petróleo estarem possivelmente associados com elevação/quedas nos anos subsequentes do número de papers sobre cada método de EOR. É possível observar ainda certa latência, ou tempo de resposta,

para um aumento no preço do petróleo acarretar em um aumento no número de trabalhos científicos acerca dos métodos de EOR, o que corrobora a influência do preço do petróleo no desenvolvimento das técnicas de EOR nas décadas supracitadas. Tal estado, influenciado pela constante e, por vezes, brusca variação do preço do petróleo, impactou fortemente a tomada de decisões para uso de EOR nas últimas duas décadas. Isso ocasionou em perdas de janelas de oportunidades, sobretudo quando se trata de aumentar a recuperação de petróleo. Contudo, especialmente a partir de 2011, nota-se que apesar da queda do preço do petróleo e de reduções temporárias no número de trabalhos científicos, o número de trabalhos científicos continuou elevado. Tal fato pode estar associado, dentre outros fatores, com: (i) o prognóstico de previsões de aumento futuro do preço do petróleo; (ii) crescentes investimentos em pesquisa na área por parte de muitas empresas operadoras que enxergam um horizonte propício para a aplicação de EOR; interesse de órgãos governamentais, vislumbrando meios de garantir segurança energética de suas nações; e surgimento de novas fronteiras exploratórias como Shale Gas e Pré-Sal que vem se mostrando como potenciais categorias de reservas para a implementação de métodos de EOR.

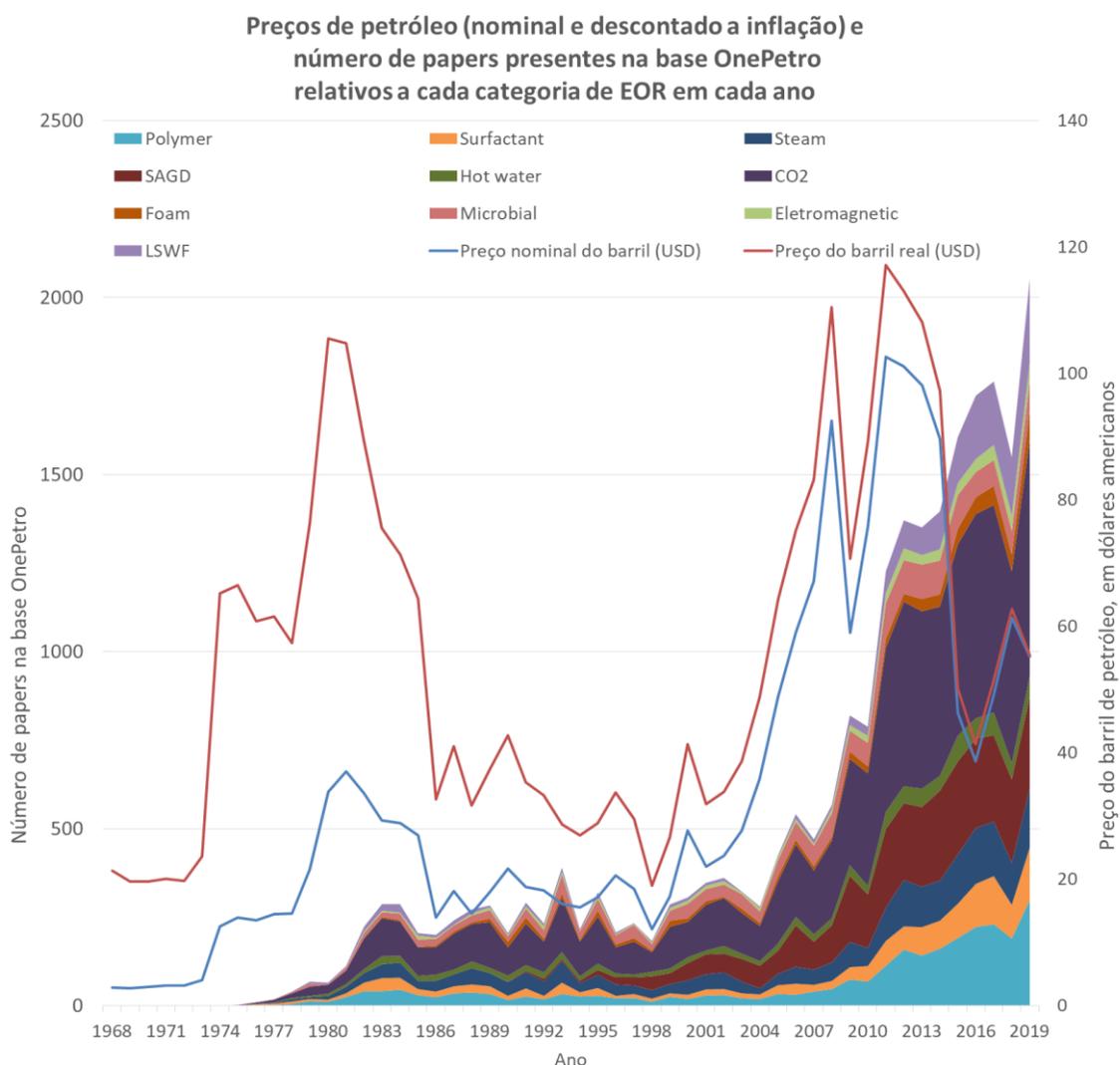


Figura 3. Preços do barril de petróleo Brent (nominal e descontada a inflação tendo como base os EUA) e número de artigos presentes na base OnePetro relativos a cada categoria de EOR ao longo do tempo. Dados do número de artigos foram obtidos a partir dos resultados de pesquisa no *OnePetro* e os dos preços de barril de petróleo são provenientes de relatório da *U.S. Energy Information Administration* (2020). O preço descontado a inflação também fora incluído para se obter uma comparação mais justa em cenários econômicos distintos^{1*}.

^{1*} Em economia, o valor nominal é medido em termos de dinheiro, enquanto o valor real é medido em relação à valoração de bens ou serviços. Um valor real é aquele que foi ajustado pela inflação, permitindo a comparação de quantidades como se os preços dos bens não tivessem mudado em média. Alterações no valor em termos reais, portanto, incluem o efeito da inflação. Em contraste com um valor real, um valor nominal não foi ajustado pela inflação e, portanto, as mudanças no valor nominal refletem pelo menos em parte o efeito da inflação. Os Preços Reais do Petróleo são calculados dividindo o preço nominal em um determinado mês pela razão do Índice de Preços ao Consumidor (IPC) naquele mês para o IPC em algum período "base". Conseqüentemente, todos os preços reais são expressos em dólares "atuais" e qualquer preço do ano atual pode ser comparado diretamente com quaisquer preços reais projetados ou passados. Dessa forma, o preço real do petróleo permite uma comparação mais justa entre projetos de EOR feitos em dados e cenários políticos-econômicos diferentes.

A utilização de EOR, atualmente, ainda é pouco expressiva globalmente, com cerca de 4 a 6% da produção mundial de petróleo (IEA, 2019). No entanto, conforme mostra o gráfico da [Figura 2](#), nas últimas décadas a pesquisa pela aplicação de EOR tem apresentado elevado crescimento. Motivados também por uma questão de segurança energética de recursos minerais, diversos países passaram a adotar incentivos fiscais para se melhorar a recuperação e estender a vida útil de reservas já depletadas, maduras ou marginais, como por exemplo nos EUA, países do Mar do Norte e Argentina. Torna-se evidente, então, o potencial de crescimento dessas aplicações, tanto com o uso de métodos de EOR já consolidados no mercado, quanto devido à inovação e criação de novas tecnologias de recuperação (MASHAYEKHIZADEH; KORD; DEJAM, 2014; TABER; MARTIN; SERIGHT, 1997).

Observa-se, atualmente, uma tendência de se avaliar possíveis implementações de métodos de EOR desde os estágios iniciais da produção. Essa tendência possibilita o desenvolvimento do campo alinhado com perspectivas de longo prazo e maximização do retorno sobre o investimento. As instalações de produção (*offshore* ou *onshore*) têm sido projetadas levando em consideração um horizonte aplicações dos métodos de EOR ao longo do ciclo de vida do campo, e demandam tecnologias e equipamentos específicos. Da mesma forma, o planejamento da aplicação de EOR o mais cedo possível permite a otimização da logística e do fornecimento de insumos futuros necessários para os diversos mecanismos de injeção no reservatório.

No início do ciclo de vida de um campo de produção as informações disponíveis acerca das características dos reservatórios costumam ser escassas, tornando muitas vezes inviável realizar estudos envolvendo simulação de reservatórios e análises detalhadas que envolvam aspectos referentes a logística e aspectos regulatórios. Além disso, muitas empresas de pequeno a médio porte por vezes não dominam o conhecimento necessário para realizar uma análise holística e breve sobre os fatores envolvidos para a aplicação do EOR. Estas empresas por vezes são operadoras, prestadoras de serviços, distribuidoras e fabricantes de uma variedade de equipamentos e produtos químicos, e outros *stakeholders*, inseridas em tal cadeia de valor, interessados

em saber sobre a viabilidade da aplicação de EOR em uma determinada região, porém limitadas pelo custo envolvido para desenvolver o estudo específico.

Ao longo da exploração e desenvolvimento de um campo cada informação adicional obtida é utilizada para se melhorar as probabilidades das projeções. Engenheiros de reservatórios estão em uma posição única, uma vez que a maioria dos reservatórios estão em profundidades de milhares de pés e suas características e comportamento não podem ser observados diretamente. Esses reservatórios só podem ser sondados e amostrados de forma muito limitada e, conseqüentemente, seus comportamentos devem ser conceituados e preditos de alguma forma. A situação é ainda mais complicada pelo fato de que muito pouco se sabe sobre um reservatório em sua descoberta, a fim de desenvolvê-lo da maneira mais eficiente. Uma grande quantidade de produção e outros dados relacionados ao reservatório só está disponível em fases posteriores, quando as reservas parcialmente produzidas e as condições termodinâmicas já não são mais iguais às iniciais. Como apontado, a recuperação final da maioria dos reservatórios de petróleo é, em geral, bastante limitada. No entanto, a coleta de dados no início do desenvolvimento de campo é enfatizada por profissionais experientes, uma vez que podem ser identificados campos semelhantes em outras regiões do mundo com desafios parecidos. Dessa forma, "lições aprendidas" no desenvolvimento de campos análogos ao redor do mundo podem ser incorporadas no processo de decisão em fases posteriores.

O sucesso da implementação de qualquer projeto de EOR é considerado dependente de uma ampla gama de variáveis. Além disso, embora os desafios relacionados ao desenvolvimento de cada campo possam apresentar características gerais de natureza similar e o uso de casos semelhantes auxilie e muito na agilidade do processo de decisão, nenhuma solução única funciona perfeitamente para todos os reservatórios (ALVARADO; MANRIQUE 2010a; MASHAYEKHIZADEH; KORD; DEJAM, 2014; SHENG, 2013). Cada reservatório é único em seu caráter e desempenho, e precisa ser avaliado individualmente. Em decorrência disso, a avaliação completa da escala de campo de um projeto EOR é uma tarefa muito dispendiosa e demorada. Dessa forma, um eficiente trabalho de seleção do melhor método EOR para um determinado reservatório

baseado em dados reais pode reduzir o custo computacional, além de economizar tempo de recursos humanos qualificados na execução e planejamento de simulações e ajustes de histórico.

Uma possível abordagem para tal processo de triagem de métodos de EOR seria por meio do estudo, caso a caso, de projetos anteriormente desenvolvidos (ALADASANI; BAI, 2010; TABER; MARTIN; SERIGHT, 1997; TRUJILLO et al., 2010). Isso demandaria a confecção de planilhas de critérios ou a construção de metodologias baseadas na ponderação de diferentes fatores para seleção do método de EOR. Estes formatos de triagem de EOR, contudo, apresentam limitações, especialmente nas seguintes condições: (i) casos pouco explorados, onde o conhecimento à priori é reduzido; (ii) quando o número de variáveis é grande e suas relações são altamente complexas; e (iii) quando a dinâmica dos reservatórios muda significativamente ao longo da produção de um campo, exigindo constantes atualizações.

Conforme ilustram (ALVARADO; MANRIQUE, 2010a), uma conclusão ingênua comum, geralmente resultante de decisões financeiras pouco estruturadas, é adiar os projetos de EOR até que o limite econômico da recuperação primária ou secundária seja atingido. Esse tipo de tomada de decisão pressupõe que condições favoráveis ou não para o processo de EOR encontradas em um reservatório, em determinado momento, prevalecerão pelo resto do ciclo de vida produtivo do campo, além de levar a perdas relativas ao intervalo específico de oportunidade. Um exemplo ilustrativo é perda de pressão que ocorre no reservatório, processo análogo ao de abrir uma garrafa de refrigerante com escape do gás e “perda” do potencial dessa energia.

Torna-se necessário, portanto, o uso de uma sistemática que permita certa flexibilidade tanto em relação às possibilidades de soluções viáveis oferecidas, como em relação a sua evolução dinâmica com o tempo. Se o processo de seleção for executado uma vez e nunca revisado à medida que os reservatórios mudam suas condições físicas e termodinâmicas, a triagem pode apresentar uma data limitada em que é válida. Dessa forma, torna-se relevante o conhecimento global dos métodos de EOR comparativamente, um com o outro, além das particularidades de cada classe de método de EOR e o

desenvolvimento de rotinas flexíveis que auxiliem na tomada de decisão acerca do desenvolvimento de um campo de petróleo.

2.3. Quanto ao uso de Bases de Dados Integradas em algoritmos de Inteligência Artificial

Os processos de tomada de decisão tanto quantitativos e qualitativos, bem como a abordagem de questões delicadas, se mostram importantes na análise e tomada de decisão de projetos de EOR. Os profissionais relativos à engenharia de reservatórios concentram-se na análise quantitativa dos mecanismos de produção e na avaliação de reservas e desempenho (simulação de reservatórios), entre muitas outras tarefas analíticas. Desse modo, a tomada de decisão se baseia em aspectos quantificáveis de um determinado problema, como o Valor Presente Líquido do projeto, para que decisões racionais possam ser tomadas. As maiores dificuldades surgem quando questões não quantificáveis se tornam parte do problema de decisão. As considerações tanto técnicas, como também sociais, ambientais, regulatórias, e logísticas, por exemplo, geralmente se apresentam como aspectos qualitativos de um problema, que podem ser difíceis de se colocar em termos quantificáveis. Para a aplicação dos métodos de EOR, fatores como fontes de matérias-primas (por exemplo, água, gás e produtos químicos para injeção), descarte de subprodutos ou resíduos, e a proximidade de fornecedores de equipamentos e insumos e de prestadores de serviço especializados, constituem um grupo de variáveis que, muitas vezes, apresentam sua quantificação limitada. Conforme (MANRIQUE, 2009), construir um modelo de decisão que se concentre nos critérios relevantes é um pré-requisito para o sucesso do modelo. Porém, outros indicadores ou formas de se avaliar devem ser utilizados para problemas difíceis de se quantificar. Dessa forma, a criação de uma metodologia de análise e predição que apresente um equilíbrio e alie variáveis qualitativas e quantitativas, é um aspecto importante e que pode ser de substancial valia.

Nesse contexto, uso de técnicas de Ciência de Dados e algoritmos de Inteligência Artificial são cruciais para integrar e analisar dados de projetos de EOR, permitindo a criação de modelos explicativos e preditivos e sistemas de

apoio à decisão. Essas ferramentas de IA permitem processar e interpretar inter-relações complexas entre variáveis diversas, facilitando a quantificação de fatores qualitativos como impactos sociais e ambientais inerentes aos projetos de EOR.

2.4. Quanto ao potencial de aplicação de EOR no Brasil

Dentre as novas fronteiras exploratórias brasileiras, a Bacia de Santos destaca-se por ser a maior bacia sedimentar *offshore* do Brasil, com área total superior a 350.000 km² que se estende de Cabo Frio (RJ) a Florianópolis (SC) (Moreira, 2007). Segundo informações do Plano de Negócios e Gestão da Petrobras para o quinquênio 2017-2021, a média de produção diária da camada Pré-sal foi de 1.456 barris por dia em janeiro de 2019, em que nos últimos 60 meses a produção média de óleo do Brasil cresceu 28%, resultado do acréscimo de 310% da produção do Pré-sal e retração de 31% da produção do Pós-sal (PETROBRAS, 2019). Dentre os campos existentes na Bacia de Santos, o campo de Libra, primeiro campo leiloadado sob a regra da partilha da produção no Pré-sal, possui um volume recuperável de óleo estimado entre 9 e 15 bilhões de barris. Apesar de Libra estar no início do desenvolvimento produtivo, já estão sendo aplicadas técnicas especiais de recuperação de petróleo. Se tomada a estimativa mais conservadora (9 bilhões), conseguindo um incremento de 1% no fator de recuperação, e considerando 60 dólares o preço médio do barril, haverá um aumento da receita bruta do consórcio em mais de USD 5 bilhões.

Somada ao potencial produtivo do Pré-sal, há também uma significativa quantidade de campos com produções declinantes ou desativados no Brasil. Estes campos ainda possuem substancial quantidade de óleo passível de ser recuperado, especialmente nas bacias sedimentares *onshore*, em sua maioria localizadas na região Nordeste do país. Com o intuito de incentivar a produção em campos maduros o governo federal brasileiro lançou o Programa de Revitalização das Atividades de Exploração e Produção de Petróleo e Gás Natural em Áreas Terrestres (REATE 2020), com incentivos fiscais para a produção em campos maduros (FAVARO, 2019). O programa visa a revitalização do setor que pode ser um importante vetor do desenvolvimento

regional, viabilizando uma indústria *onshore* competitiva, com produção crescente e com estímulo à criação de empresas nacionais, incluindo as startups (empresas de base tecnológica) e a atração de investimentos estrangeiros, levando uma pluralidade de operadores e diversidade de fornecedores de produtos e serviços (MINISTÉRIO DE MINAS E ENERGIA, 2019). Além disso, e diante da necessidade de revitalizar a produção em campos maduros, o CNPE (Conselho Nacional de Política Energética), por meio da resolução 17/2017 delegou à ANP conceder redução de royalties, de 10% para até 5%, sobre a produção incremental dos campos - mediante pedido da empresa operadora e comprovado o benefício econômico para os entes federados (DELGADO, 2019).

2.5. Quanto à interdisciplinaridade na aplicação de EOR

O estudo e a aplicação de tópicos relacionados aos métodos de EOR costumam envolver diversas áreas da ciência, como ilustra a Figura 4. Além disso, conforme afirma (SHENG, 2013), o planejamento, condução e a execução de um projeto de EOR requer especialistas e experiência em diferentes disciplinas. De forma realista, um grupo de especialistas em métodos de EOR pode não estar prontamente disponível em uma companhia, e ainda que esse grupo de especialistas esteja disponível, eles podem ter experiência limitada em enfrentar desafios especiais em um projeto específico. Como o desenvolvimento de projetos de EOR costuma envolver custos da ordem de dezenas milhões de dólares ou mais, a experiência necessária pode não ser totalmente obtida via livros didáticos, podendo ser interessante uma abordagem suportada pela experiência de projetos de campo reais. Além disso, uma consideração adicional importante na tomada de decisão relativa ao desenvolvimento de campos de petróleo é o viés cognitivo, que pode assumir várias formas, uma das quais representando as limitações cognitivas da mente humana (WELSH et al., 2005; BEGG et al., 2003). Para mitigar tal viés cognitivo, existem diversas abordagens, dentre as quais a baseada em dados coletados acerca da implementação de projetos de EOR no mundo. Nesse contexto, estratégias orientadas por dados podem ser utilizadas como parte do processo de suporte à tomada de decisão. Dessa forma, ao invés de depender somente do conhecimento e das

experiências sobre os campos de produção advindos de apenas alguns especialistas, inúmeros conhecimentos e diversas experiências podem ser incorporados ao problema de decisão.

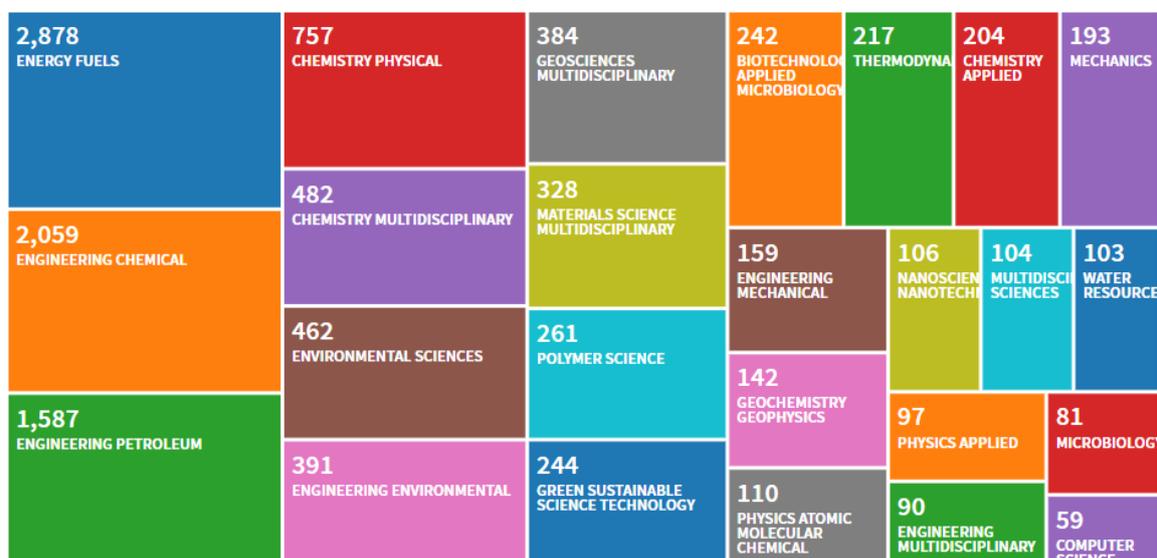


Figura 4. Número de registro de trabalhos relacionados ao tópico "Enhanced Oil Recovery" na base Web of Science, para cada categoria. Verifica-se uma grande interdisciplinaridade em tópicos de pesquisa relacionados ao EOR.

Dada a relevância dos métodos de EOR como tecnologia para se potencializar a produção de petróleo, e da IA como uma vantagem estratégica na análise e resolução de problemas complexos, é então pertinente a convergência de ambas as técnicas na avaliação de amplos bancos de dados de EOR por meio de novas perspectivas. Essa integração proporciona a investigação das relações de diversas variáveis com a eficiência dos tratamentos de EOR. Além disso, com o advento de uma sistemática integrativa, torna-se viável o suporte à decisão na produção de campos de petróleo de interesse, tais como os da bacia de Santos e de campos (Pré-Sal e Pós-Sal), cuja aplicação de métodos EOR apresenta-se em latentes perspectivas de maior aproveitamento de suas reservas.

2.6. Quanto à importância de uma análise mundial da aplicação dos métodos de EOR

O avanço dos métodos de EOR se mostra como uma das tecnologias de maior potencial para se atacar o problema de baixos fatores de recuperação, otimização da produção de campos maduros e de auxílio em políticas visando a segurança energética de diversas nações. Nesse contexto, para se ter uma visão

melhor dessa problemática é essencial entender como os métodos de EOR se desenvolveram ao longo das décadas em diversas regiões do globo, e quais são os principais limitantes de sua difusão e aplicação. Tal dinâmica tem se mostrado dependente de um grande número de fatores e sua análise pode se tornar uma importante forma de subsidiar tomadas de decisão nesse campo.

A análise de variáveis petrofísicas se mostra necessária para o entendimento mais profundo dos principais fatores técnicos, e em que tipos de ativos, que alavancam ou limitam o desenvolvimento de determinado método de EOR, bem como desafios tecnológicos a serem superados.

A análise da influência do preço do barril, aspectos comerciais e logísticos locais possibilitam auxiliar na compreensão de fatores externos que podem exercer influência na continuidade da aplicação de EOR em diversos casos.

A análise dos perfis e comportamento de empresas de diversos portes frente e suas relações com diferentes tipos de ativos e graus de depleção e diferentes métodos de EOR pode possibilitar compreender com maior profundidade o posicionamento estratégico de diferentes companhias em diversas situações. Somado a isso, a análise de como diferentes países promoveram a adoção de técnicas de EOR por meio de marcos regulatórios pode trazer maior entendimento dos fatores que podem estimular a aplicação de técnicas de EOR por diferentes empresas, a revitalização de campos maduros e consequente desenvolvimento de política de segurança energética.

2.7. Quanto à importância de obtenção de campos análogos

As descobertas de novas reservas de petróleo e atividades exploratórias encontram-se em taxas declinantes. A capacidade de manter a produção em diversas regiões do globo tem sido uma preocupação recorrente nos últimos 50 anos (KANG; LIM; HUH, 2016; LE et al., 2015). Dentre as formas de se aumentar a competitividade da indústria e, principalmente, suprir o abastecimento de petróleo para as próximas décadas, está a exploração de campos maduros. National Oil Companies (NOC) e International Oil Companies (IOC) têm realizado um movimento de vender seus ativos maduros, com reservas parcialmente

depletadas, para operadoras de pequeno e médio porte com o intuito de focar em ativos globais de grande porte, com maiores taxas internas de retorno no curto prazo. Nesse contexto, a aquisição desses ativos por parte de empresas de menor porte requer, em sua maioria, a elaboração de novos planos de desenvolvimento, incluindo novas estratégias de produção, técnicas de recuperação melhorada e avançada.

A identificação de campos análogos se mostra como uma ferramenta complementar na avaliação técnica e estratégica do ativo de forma mais célere, especialmente se comparada a avaliações econômicas tradicionais e simulações de reservatório. O Petroleum Resources Management System (PRMS) define reservatórios análogos como aqueles que possuem propriedades semelhantes de (i) rocha e fluido; (ii) condições de reservatório (profundidade, temperatura e pressão) e de mecanismos de produção; (iii) se encontrem em um estágio mais avançado de desenvolvimento em relação ao reservatório de interesse e que possibilitem auxiliar nos planos de exploração e desenvolvimento, na avaliação de estratégias de produção e estimativas de recuperação.

Encontrar análogos adequados de um reservatório é um processo complexo, envolvendo a correspondência de variáveis de diversas disciplinas (propriedades geológicas e petrofísicas, de fluido e de engenharia). Os fluxos de trabalho tradicionais para identificação de campos análogos dependem do julgamento do avaliador e são restritos à sua experiência, carecendo de uma abordagem mais sistemática e objetiva. É comum que geocientistas confiem no uso de apenas um único análogo, normalmente de formações geológicas vizinhas ao reservatório de interesse, gerando um espectro restrito e enviesado de avaliação. Trabalhos recentes de obtenção de campos análogos incluem o uso abrangente de informações de projetos de exploração e produção, trazendo análises comparativas do desempenho e das estratégias de recuperação utilizadas (JIA et al., 2016; LU, 2021; SUN et al., 2021; SUN; POLLITT, 2021). Outros trabalhos inovam ao trazerem abordagens orientadas por dados; porém, em geral, são limitados devido à abrangência de dados, seja em relação à quantidade e distribuição dos projetos tomados como espaço de busca ou a conjuntos limitados de variáveis consideradas (GALICIA; REYES;

SHEREMETOV, 2021; MUMTAZ et al., 2020; OLUKOGA; FENG, 2021; PRATAMA, 2021; RODRIGUEZ et al., 2013; TEMIZEL; DURSUN, 2013).

Nesse sentido verifica-se a oportunidade e demanda para desenvolvimento de um procedimento sistemático para a busca de reservatórios análogos com base num abrangente banco de dados mundial, subsidiado por técnicas de *machine learning* e validada por análises comparativas. Tal abordagem pode fornecer flexibilidade para avaliar diferentes cenários e prioridades em relação às variáveis de maior interesse, além de viabilizar comparações e aprendizado com experiências internacionais. Dessa forma, a sistematização da busca por reservatórios análogos, usados em conjunto com os dados primários, podem expandir o conhecimento do indivíduo e da equipe, e desenvolver percepções que não seriam viáveis com o uso de dados locais ou da experiência individual isoladamente.

3. FUNDAMENTOS TEÓRICOS

3.1. Mineração de dados

Mineração de dados é o processo de explorar dados à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, desenvolvendo assim novos subconjuntos de dados. Tal processo consiste em um tópico recente em ciência da computação, e que utiliza várias técnicas da estatística, aquisição de informação, inteligência artificial e reconhecimento de padrões (AGGARWAL, 2015; SKIENA, 2017).

As técnicas da Mineração de Dados podem ser categorizadas como preditivas ou descritivas. O objetivo das técnicas preditivas é prever os valores desconhecidos de uma variável alvo com base em dados conhecidos. O objetivo das técnicas descritivas é identificar padrões interpretáveis por humanos e associações em dados, por exemplo, agrupamento e regras de associação. A descoberta metódica de inter-relações e padrões úteis nos dados é possibilitada por um conjunto de etapas iterativas conhecidas coletivamente como processo de ciência de dados. Ao longo dos anos de evolução da ciência de dados, diferentes estruturas para o processo de mineração de dados e extração do conhecimento foram apresentadas por vários órgãos acadêmicos e comerciais, como por exemplo Knowledge-Discovery in Databases (KDD), Cross Industry Standard Process for Data Mining (CRISP-DM) e Sample, Explore, Modify, Model, and Assess (SEMMA) (FAYYAD, 1996; LTIFI et al., 2013; MARTINEZ-PLUMED et al., 2020). Tais abordagens e suas adaptações têm se mostrado relevantes em diferentes contextos, apresentando uma grande diversidade de aplicações (DIAMANTINI; POTENA; STORTI, 2013; GONZÁLEZ-CALABOZO; VALVERDE-ALBACETE; PELÁEZ-MORENO, 2016; KADI; IDRI; FERNANDEZ-ALEMAN, 2017a; LOTFNEZHAD AFSHAR et al., 2015; NAGASHIMA; KATO, 2019; RAMOS et al., 2017; ROGALEWICZ; SIKI, 2016). Além disso, a possibilidade de se integrar técnicas de machine learning a adaptações de modelos como KDD e CRISP-DM viabiliza tornar mais robustas análises e predições em conjuntos de dados com grandes dimensões (HUBER et al., 2019;

KADI; IDRI; FERNANDEZ-ALEMAN, 2017b; STUDER et al., 2021; TRIPATHI et al., 2020; WIEMER; DROWATZKY; IHLENFELDT, 2019).

3.2. Técnicas de visualização de bases de dados de alta dimensionalidade

Um possível caminho para a extração do conhecimento a partir dos dados, especialmente onde se requer maiores esforços cognitivos para compreensão dos fatores envolvidos, são os pipelines de visualização de dados (Sacha et al., 2014). As visualizações usam os dados ou modelos gerados a partir dos dados e permitem que os analistas detectem relacionamentos nos dados. Dada a capacidade hoje sem precedentes para adquirir, processar e integrar dados de diversas fontes e naturezas, técnicas de visualização de dados multidimensionais mostram-se uma ferramenta poderosa para analisar grandes conjuntos de dados e extrair padrões e tendências (LIU et al., 2017; STEFANOVIČ; KURASOVA, 2011). Dentre formas de amplo emprego atual de visualizações de dados diretas e tradicionais que possuem aplicações em conjuntos de dados com número alto de dimensões estão *boxplots*, gráficos de dispersão, matrizes de correlação, *treemaps*, dendrogramas e plotagens de coordenadas paralelas, dentre outras, em que diversas linguagens de programação possuem suporte de uma ou mais dessas visualizações (ITOH et al., 2017; LI; HOU, 2017; RAGHAV et al., 2016). Além disso, muitas dessas oferecem a possibilidade de serem interativas, o que viabiliza um processo mais ágil de exploração dos dados por parte do usuário (HIRVE; PRADEEP REDDY, 2019; QIN, 2019). Como forma de visualização indireta, há os métodos de redução de dimensionalidade que permitem transformar os vetores amostra, correspondentes aos dados multidimensionais, em um espaço de dimensão inferior, dentre os quais os métodos mais populares estão a análise de *components* principais e a análise discriminante linear. Contudo, tais técnicas não incorporam em seus processos relações não lineares entre os dados e possuem dificuldades para lidar com dados de diferentes formatos e naturezas.

3.3. Self-Organizing Maps

No âmbito das Redes Neurais Artificiais, os Self-Organizing Maps (SOM) tratam de uma ferramenta bioinspirada de análise e visualização de dados, com abordagem não-supervisionada, capaz de estabelecer relações topológicas e topográficas projetadas em um mapa, que se baseiam na utilização de quantização vetorial em um espaço n-dimensional e medidas do vetor similaridade. Os SOMs mimetizam computacionalmente funções semelhantes às redes neurais biológicas, como a estrutura interna dos sinais de entrada, a coordenação das atividades da unidade através das conexões laterais entre as unidades e a especialização de determinados conjuntos de neurônios com determinados agrupamentos de dados (KOHONEN, 2013). O espaço n-dimensional está relacionado ao número de variáveis envolvidas e a quantização vetorial trata-se de uma extensão da quantização escalar para um espaço multidimensional, onde os vetores, caracterizados não apenas pelos valores, mas pela orientação no espaço constituem a base analisada.

O método SOM tem sido utilizado em diferentes aplicações recentes. O algoritmo SOM destaca-se por apresentar uma abordagem guiada pelo dado e demonstra três principais vantagens, a saber: (i) identificação de relações de similaridade entre agrupamentos de dados; (ii) análise visual de fácil compreensão na forma de mapas do espaço multidimensional estudado; e (iii) identificação e individualização de padrões, observando constituição, geometria e abrangência das classes geradas, o que a torna eficiente na representação de conjuntos de dados de alta dimensionalidade (GARDNER et al., 2019; MIKAEIL, 2018; NAKAGAWA, 2019; QIAN, 2019; SEGEV, 2012; SEGEV; KANTOLA, 2012; YANG; LEE; WU, 2018; YOKOTA et al., 2020).

A técnica SOM pode ser vista como um tecido neural artificial elástico de neurônios ajustados ao espaço do sinal de input, preservando as relações topológicas do espaço do sinal (KOHONEN, 2001). A Tabela 1 sintetiza os princípios básicos envolvidos no algoritmo SOM. O espaço do sinal de input é o conjunto de vetores N-dimensionais $x_i = [x_{i1}, x_{i2}, \dots, x_{in}, \dots, x_{iN}]$ para $i = 1, 2, 3, \dots, M$ que representam a base de dados integrada X de tamanho $M \times N$. A camada de saída é uma coleção ordenada de neurônios que normalmente são

organizados como redes hexagonais ou retangulares. Cada neurônio é conectado à camada de entrada por meio de vetores peso w_i , possuindo, portanto, duas camadas. A dimensionalidade do vetor peso de cada neurônio de saída é a mesma que a dimensionalidade do vetor de entrada, portanto $w_i = [w_{j1}, w_{j2}, \dots, w_{jn}, \dots, w_{jN}]$ para $j = 1, 2, 3, \dots, J$, onde J é o número total de neurônios no mapa. Esses vetores de peso dos neurônios de saída são comparados com os vetores de entrada de acordo com algumas medidas de distância, a fim de determinar o grau de ativação daquele neurônio. A distância euclidiana é geralmente usada como critério para comparar o vetor de entrada com os vetores peso dos neurônios de saída. Após a comparação, o neurônio cujo vetor de referência possui maior similaridade com x_i é identificado e definido como o neurônio vencedor, $b_i = \text{argmin} \|x_i - w_i\|$ para $i = 1, 2, 3, \dots, M$ e $j = 1, 2, 3, \dots, J$, que é chamado de unidade de melhor ajuste (BMU) para a entrada.

Após a seleção do neurônio vencedor b_i , o vetor de peso do neurônio vencedor é atualizado para torná-lo mais próximo do vetor de entrada. Além disso, os vetores de peso dos neurônios vizinhos também são atualizados, mas em menor grau. O parâmetro da taxa de aprendizagem e a função de influência na vizinhança podem ser definidos como diminuindo gradualmente durante o processo de treinamento. Como exemplo de função de vizinhança, é comumente utilizada a Gaussiana, dada por

$$h_{bj} = \exp \left(- \frac{\|r_b - r_j\|^2}{\sigma(t)} \right)$$

onde r_i é a posição dos neurônios no mapa hexagonal dos neurônios de saída; r_j é a localização do j -ésimo neurônio, com $r_j \in R^2$; r_b é a posição do neurônio vencedor no mapa de saída; $\sigma(t)$ o alcance de vizinhança da influência dos neurônios, que reduz monotonicamente com o tempo a fim de garantir a convergência. Tal função de vizinhança é incorporada à função de atualização de peso dada por:

$$w_j(t+1) = w_j(t) + \alpha(t) h_{bj}(t) \|x_i(t) - w_j(t)\|$$

O processo adaptativo modifica os pesos da rede à medida em que auto-organização da rede é atingida. A fase de treinamento termina quando o número de épocas excede o número total predeterminado de iterações, e os neurônios no mapa de saída serão rotulados com seus nomes de entrada correspondentes.

Dois parâmetros a serem especificados e que influenciam na acurácia do SOM são o número de neurônios no tecido neural e a razão entre as dimensões de altura e largura do mapa. Quanto maior o tamanho do mapa, menor será o erro de quantização, mas maior será o erro topográfico, que é relacionado à projeção dos BMUs, e maior será o custo computacional. Por vezes, o conjunto de dados pode conter apenas alguns conjuntos de padrões de interesse a serem observados, em que uma resolução menor é suficiente. No entanto, se houver a suspeita de que há estruturas interessantes nos dados, um tecido neural artificial maior é necessário para se obter uma resolução suficiente para a investigação. As matrizes SOM típicas variam de algumas dezenas a algumas centenas de nós, e a relação das dimensões horizontal e vertical da matriz deve cumprir pelo menos aproximadamente com a relação das duas maiores *components* principais dos dados de entrada, respectivamente. Não é possível adivinhar ou estimar o tamanho exato mais adequado da matriz de antemão, devendo ser determinado pelo método de tentativa e erro, com checagem da qualidade para se obter uma solução ótima razoável. Portanto, um compromisso deve ser feito entre resolução e precisão estatística (KOHONEN, 2013).

Existem diversos métodos para visualização do espaço topológico criado no ambiente SOM. Dentre eles, os métodos mais utilizados são a matriz unificada de distâncias (U-Matrix) e o mapa de *components* e o mapa de histograma de acertos. A matriz de distância unificada é dada por

$$\text{U-Matrix} = \begin{pmatrix} u_{11} & u_{11|12} & u_{12} & u_{12|13} & \cdots & u_{1y} & u_{11|21} & u_{11|22} & \cdots & u_{1y|2y} & \vdots & \vdots & \vdots \\ u_{x1} & u_{x1|x2} & u_{x2} & u_{x2|x3} & \cdots & u_{xy} & & & & & & & \end{pmatrix}$$

e reúne as distâncias euclidianas médias do vetor de peso de um neurônio aos vetores de peso de seus neurônios vizinhos. Os valores dos elementos $u_{ij|i(j+1)}$ são as distâncias entre os neurônios vizinhos na mesma linha e coluna na grade hexagonal do mapa e os valores dos elementos u_{ij} a as médias dos elementos vizinhos da U-Matrix. Na paleta de cores utilizada no código dos resultados desse trabalho, cores mais próximas ao azul representam uma similaridade maior e as cores mais próximas do vermelho uma maior

dissimilaridade. Dessa forma, a área com valores baixos na matriz U forma o cluster, enquanto os valores mais altos da matriz indicam a fronteira existente entre dois o mais *clusters*. As distâncias $u_{ij|i(j)}$ presentes nas diagonais da matriz são nulas e na plotagem da U-Matrix representam os próprios vetores BMUS.

As *component plots* mostram a distribuição de diferentes *components*, variáveis e propriedades no espaço topológico do SOM, em que cada *component plots* está associada a uma variável original. O posicionamento topológico dos BMUs, que representam conjuntos dos vetores amostra, é igual para todas as *component plots* geradas a partir da projeção da rede treinada. Tal fato faz com que mesmas amostras/reservatórios se localizem nas mesmas posições nas diferentes *component plots*, o que torna muito fácil visualizar e pesquisar padrões de interesse entre as diferentes variáveis.

Tabela 1. Pseudocódigo do algoritmo *Self-Organizing Maps* (SOM), explicitando os principais parâmetros de entrada, processo de inicialização, treinamento e aprendizagem. (Autoria própria)

- (0) Definição de parâmetros:
 - formato de rede (uni, bi ou tridimensional)
 - geometria de vizinhança (quadrada, hexagonal, triangular/icosaedro)
 - forma da hiper superfície (planar, cilíndrica ou toroidal)
 - tamanho do mapa (altura e largura)
 - taxa inicial de aprendizado
 - número de iterações T
- (1) **input:** conjunto de vetores $x_i = [x_{i1}, x_{i2}, \dots, x_{in}, \dots, x_{iN}]$, que formam a base de dados integrada X
- (2) **output:** conjunto de vetores peso $w_i = [w_{j1}, w_{j2}, \dots, w_{jn}, \dots, w_{jN}]$
início do processo de treinamento
- (3) inicializa $w_i = [w_{j1}, w_{j2}, \dots, w_{jn}, \dots, w_{jN}]$ randomicamente
- (4) **repetir** passos (5) a (9) até $t < T$
- (5) para cada vetor $x_i \in X$,
- (6) encontrar o neurônio de melhor ajuste:

$$b_i = \text{argmin} \|x_i - w_i\|$$

atualizar o peso do BMU:

$$(7) \quad w_j(t+1) = w_j(t) + \alpha(t)h_{bj}(t) \|x_i(t) - w_j(t)\|$$

reduzir o raio de vizinhança topológica:

$$(8) \quad h_{bj} = \exp\left(-\frac{\|r_b - r_j\|^2}{(t)}\right)$$

atualizar a taxa de aprendizagem:

$$(9) \quad \alpha(t) = \alpha_0 \left(1 - \frac{t}{T}\right)$$

(10) fim do processo de treinamento

composição da U-Matrix:

$$(11) \quad \text{U-Matrix} = \begin{pmatrix} u_{11} & u_{11|12} & u_{12} & u_{12|13} & \cdots & u_{1y} & u_{11|21} & u_{11|22} & \cdots & u_{1y|2y} & \vdots & \vdots & \vdots \\ \vdots & u_{x1} & u_{x1|x2} & u_{x2} & u_{x2|x3} & u_{xy} & & & & & & & \end{pmatrix}$$

3.4. Trabalhos prévios envolvendo análise de projetos de EOR

No decorrer da última década inúmeros trabalhos científicos foram apresentados propondo análises da implementação de projetos de EOR em escala global e metodologias que fundamentassem a escolha dos métodos de EOR compatível com um tipo específico de reservatório. A investigação de amplo conjunto de aplicações de EOR pode trazer respostas para questões relevantes ao desenvolvimento de campos maduros. No decorrer da última década, inúmeros trabalhos científicos desenvolveram análises da implementação de projetos de EOR em escala global. Alguns destes trabalhos abordaram metodologias que fundamentassem a escolha dos métodos de EOR compatível com um tipo específico de reservatório. Os trabalhos de (BABADAGLI, 2007, 2020; MANRIQUE, 2009; MUGGERIDGE et al., 2014; THOMAS, 2008) abordam aplicações de EOR de forma mais genérica. Já os trabalhos (BELHAJ et al., 2020; FAKHER; IMQAM, 2020; GBADAMOSI et al., 2018, 2019; OLAJIRE, 2014; ZHANG; WEI; BAI, 2018) abordam alguma subclasse de método de recuperação em específico de forma mais aprofundada ou possuem ênfase na aplicação de EOR em alguma determinada região do mundo. Tais trabalhos anteriores, a

partir de experiências prévias de EOR, trazem detalhes e lições relevantes para a indústria acerca da implementação de EOR. No entanto, a análise e interpretação de tais dados de maneira integrada permanece um desafio significativo. A visualização e extração de informações relevantes e significativas a partir de bases de dados com número de dimensões elevado é por vezes notoriamente complexa e complicada (BERTINI; TATU; KEIM, 2011). Além disso, a abrangência dos dados utilizados por esses trabalhos como base por vezes é limitada, seja em relação aos métodos de EOR e às regiões do mundo que contemplam, ou mesmo em relação às variáveis elencadas para análise ou à diversidade de fontes utilizadas. Verifica-se ainda que há uma lacuna na literatura científica e tecnológica quanto a métodos de visualizações multidimensionais dos dados e a integração de diferentes formas e fontes de dados no âmbito de recuperação avançada de petróleo.

Dentre os trabalhos encontrados, os que receberam mais citações ao longo do tempo por parte da comunidade científica e/ou mais recentes são apresentados nas figuras 5, 6 e 7 a seguir. As figuras 5 e 6 condensam, respectivamente, informações a respeito dos métodos de EOR e das variáveis consideradas para avaliar a aplicação dos métodos de EOR. A Figura 7 sintetiza informações quanto à abrangência da bases de dados utilizada e a que métodos de visualização de dados e análise foram utilizados no trabalho.

Artigo	Revista	Ano	Citações (Google Scholar)	Média de citações por ano desde publicação (↓)	Térmicos																	Injeção de gases										Químicos						Qtd de métodos de EOR	Qtd de variáveis	Métodos EOR + variáveis
					Low Salinity Waterflooding	Vapor	Água quente	Combustão	SAGD	Vapor cíclico	Hidrocarb. misc.	Hidrocarb.imisc.	Gás pobre	Gás rico	Gás ácido misc.	CO2 miscible	CO2 imiscível	Nitrog. miscível	Nitrogen imiscível	Gás de comb. misc.	Gás de comb. imisc.	Nitratos/Azotos	Air	FAWAG	SWAG	WAG	Polímero	Surfactante	Polímero & Surfactante	Alcaline	Alcaline & Surfactante	Alka. Surf & Polímero	Microbial	Partículas termicamente ativadas (TAP)	Mineração de superfície	Aquecimento eletromagnético	Acústico			
Alvarado & Manrique (2010)	Energies	2010	728	66,18		X	X	X	X	X	X	X	X	X	X					X			X	X		X	X	X	X						20	15	35			
Standness & Skjevrak (2014)	Journal of Petroleum Science and Engineering	2014	133	19,00																			X											1	14	15				
Jiang et al. (2019)	Applied Energy	2019	37	18,50								X	X																					2	10	12				
Adasani (2011)	Journal of Petroleum Science and Engineering	2011	151	15,10		X	X	X		X	X		X	X	X	X				X	X	X	X	X	X	X	X	X	X						19	10	29			
Zhang et al. (2018)	Fuel	2018	31	10,33										X																			1	11	12					
Mokheimer et al. (2019)	Journal of Energy Resources Technology	2019	20	10,00		X	X	X	X	X																								5	10	15				
Saleh et al. (2014)	SPE Reservoir Evaluation & Engineering	2014	59	8,43																			X											1	10	11				
Kamari et al. (2014)	Neural Computing and Applications	2014	51	7,29		X	X	X		X			X	X													X							7	7	14				
Gao (2011)	Journal of Petroleum Exploration and Production Technology	2011	65	6,50																			X											1	6	7				
Liu et al. (2020)	Journal of Petroleum Science and Engineering	2020	6	6,00	X	X	X	X	X	X			X	X	X	X				X		X	X	X	X	X								16	9	25				
Gbadamosi et al. (2018)	Journal of Petroleum Exploration and Production Technology	2018	17	5,67																		X	X	X										4	11	15				
Gao & Kekri (2011)	Energy Sources, Part A: Recovery, Utilization, and Environmental Effects	2011	46	4,60																							X							1	7	8				
Gao et al. (2013)	Greenhouse Gases: Science and Technology	2013	32	4,00								X	X																					2	10	12				
Zhang et al. (2016)	SPE Improved Oil Recovery Conference	2016	17	3,40																			X											1	14	15				
Siena et al. (2016)	SPE Reservoir Evaluation & Engineering	2016	10	2,00		X	X	X		X	X		X	X	X							X	X	X										11	6	17				
Faker & Imqam (2020)	Springer Nature Applied Sciences	2020	2	2,00																														0	4	4				
Saleh et al. (2017)	SPE Reservoir Evaluation & Engineering	2017	5	1,25																			X											1	9	10				

Figura 5. Métodos de EOR abordados pelos principais artigos revisados, que possuem resultados na análise de projetos de EOR. São listadas as revistas e ano em que foram publicados, números de citações obtidos no Google Scholar, com ordenação decrescente pelo número médio de citações por ano para cada artigo. Além disso, também é listada uma contabilização com o número de métodos de EOR e de variáveis que cada trabalho aborda, dividindo-se nas suas principais classes, a saber: métodos térmicos, injeção de gases, injeção de produtos químicos, e outros.

Artigo	Revista	Citações (Google Scholar)	Média de citações por ano desde publicação (↓)	Prop. Petrofísicas Reservat.							Prop. óleo e água formação							Configurações de poços e variáveis de projeto												Qtd de variáveis	Qtd de métodos de EOR	Variáveis + métodos EOR													
				Litologia	Compressibilidade	Profundidade	Temperatura	Espessura payzone	Mergulho da formação	Porosidade	Permeabilidade média	Varição de permeabilidade	Dykstra Parsons	°API	Viscosidade	Composição	MMP	Saturação inicial de óleo	Saturação final de óleo	Óleo restante in place (ROIP)	Teor de argila	Água formação	Salinidade	Concentração de divalentes	Mobabilidade	Pressão mínima de miscibilidade	Pressão atual	Pressão inicial	Fratura				Área (acres)	Well pattern	Espaçamento	N° poços	Tipo de polímero injetado	Concentração injetada	Slug pore size (PV)	Viscosidade fluido injetado (polímero)	Peso molecular do polímero	Water cut	Aumento da produção	Fator de recuperação	Produção incremental de óleo por fluido injetado
Alvarado & Manrique (2010)	Energies	728	66,18	X	X	X	X	X	X		X	X				X										X											X	X	X	X	X	15	20	35	
Standness & Skjevraak (2014)	Journal of Petroleum Science and Engineering	133	19,00	X		X			X	X									X									X	X	X	X							X			14	1	15		
Jiang et al. (2019)	Applied Energy	37	18,50			X	X		X	X		X	X		X											X		X								X				10	2	12			
Adasani (2011)	Journal of Petroleum Science and Engineering	151	15,10	X	X	X			X	X		X	X		X																				X					10	19	29			
Zhang et al. (2018)	Fuel	31	10,33	X	X	X	X		X	X		X	X	X															X						X					11	1	12			
Mokheimer et al. (2019)	Journal of Energy Resources Technology	20	10,00			X		X	X		X	X		X										X	X										X	X				10	5	15			
Saleh et al. (2014)	SPE Reservoir Evaluation & Engineering	59	8,43			X			X	X		X							X									X	X	X	X	X							10	1	11				
Kamari et al. (2014)	Neural Computing and Applications	51	7,29			X	X		X	X		X	X		X																									7	7	14			
Gao (2011)	Journal of Petroleum Exploration and Production Technology	65	6,50			X			X	X		X	X																					X					6	1	7				
Liu et al. (2020)	Journal of Petroleum Science and Engineering	5	5,00			X	X		X	X		X	X						X															X					9	16	25				
Gbadamosi et al. (2018)	Journal of Petroleum Exploration and Production Technology	17	5,67	X	X	X	X		X		X	X											X											X					11	4	15				
Gao & Kekri (2011)	Energy Sources, Part A: Recovery, Utilization, and Environmental Effects	46	4,60			X			X		X																	X						X	X				7	1	8				
Gao et al. (2013)	Greenhouse Gases: Science and Technology	32	4,00			X	X		X	X		X	X											X	X			X												10	2	12			
Zhang et al. (2016)	SPE Improved Oil Recovery Conference	17	3,40	X		X			X	X		X							X	X						X	X	X	X	X	X	X							14	1	15				
Siena et al. (2016)	SPE Reservoir Evaluation & Engineering	10	2,00			X	X		X	X		X	X																											6	11	17			
Faker & Imqam (2020)	Springer Nature Applied Sciences	1	1,00			X					X	X																X											4	0	4				
Saleh et al. (2017)	SPE Reservoir Evaluation & Engineering	5	1,25						X	X		X							X	X								X	X	X	X									9	1	10			

Figura 6. As variáveis abordadas pelos principais artigos revisados, que possuem resultados na análise de projetos de EOR. Para cada artigo são listadas as revistas e ano em que foram publicados, números de citações obtidos no Google Scholar, com ordenação decrescente pelo número médio de citações por ano. Além disso, também é listada uma contabilização com o número de métodos de EOR e de variáveis que cada trabalho aborda.

Artigo	Revista	Ano	Citações (Google Scholar)	Média de citações por ano desde publicação (↓)	Quantidade de projetos	Região analisada	Análise descritiva	Análise qualitativa	Tabelas comparativas	Gráfico de tendência	Barras	Pizza	Scatterplots	Histograma	Boxplots
Avarado & Manrique (2010)	Energies	2010	728	66,18	1507 projetos de EOR	Mundo	X	X	X	X	X	X			
Standness & Skjevrak (2014)	Journal of Petroleum Science and Engineering	2014	133	19,00	72 projetos de inj. polímero	EUA, Canada, China, Alemanha, Brasil, UK, Índia, França, Suriname, Angola e Oman	X		X		X		X		
Jiang et al. (2019)	Applied Energy	2019	37	18,50	40 projetos inj. CO2	EUA	X			X			X	X	
Adasani (2011)	Journal of Petroleum Science and Engineering	2011	151	15,10	652 projetos de EOR	EUA, Canada, China, Noruega, Brasil, UK, Venezuela, Alemanha, Argentina, Colômbia, Indonésia, Turquia, UAE, Líbia, Dinamarca, Romênia, Trindade	X	X		X	X	X		X	X
Zhang et al. (2018)	Fuel	2018	31	10,33	41 projetos de inj. CO2	EUA, China, Turquia, Trindade, Malásia, Hungria, Argentina, Canada e Brasil	X				X	X	X	X	X
Mokheimer et al. (2019)	Journal of Energy Resources Technology	2019	20	10,00	25 projetos térmicas	EUA, Canada, Venezuela, China, Brasil, Índia, Romênia, Oman e Turquia	X	X							
Saleh et al. (2014)	SPE Reservoir Evaluation & Engineering	2014	59	8,43	250 projetos de inj. polímero	China, EUA, Canada, Índia, Alemanha, Argentina e França						X	X	X	X
Kamari et al. (2014)	Neural Computing and Applications	2014	51	7,29	Projetos de EOR no mundo. Não menciona a quantidade.	Não menciona localidade	X						X		
Gao (2011)	Journal of Petroleum Exploration and Production Technology	2011	65	6,50	10 projetos de aplicação microbiana	EUA, China, Argentina e Malásia	X								
Liu et al. (2020)	Journal of Petroleum Science and Engineering	2020	6	6,00	120 projetos de EOR	Mundo	X	X	X		X	X			
Gbadamosi et al. (2018)	Journal of Petroleum Exploration and Production Technology	2018	17	5,67	20 campos de injeção WAG	Mar do Norte (Noruega, UK, Alemanha)	X	X	X				X		
Gao & Kekri (2011)	Energy Sources, Part A Recovery, Utilization, and Environmental Effects	2011	46	4,60	10 projetos de injeção microbiana	Malásia, EUA, Argentina e China	X		X		X	X			
Gao et al. (2013)	Greenhouse Gases: Science and Technology	2013	32	4,00	5 projetos de inj. CO2	EUA, Turquia, Trindade e China	X	X	X		X				
Zhang et al. (2016)	SPE Improved Oil Recovery Conference	2016	17	3,40	55 projetos de inj. polímero	China					X	X	X		X
Siena et al. (2016)	SPE Reservoir Evaluation & Engineering	2016	10	2,00	250 projetos de EOR	EUA, Canada, China, Brasil, Noruega, UK, Argentina, Indonésia, Angola, França							X		
Faker & Imqam (2020)	Springer Nature Applied Sciences	2020	2	2,00	200 testes de laboratório e 20 projetos de campo (inj. CO2)	Não menciona	X					X	X	X	X
Saleh et al. (2017)	SPE Reservoir Evaluation & Engineering	2017	5	1,25	70 projetos de inj. polímero	China, EUA, Canada, Brasil, Alemanha, Argentina, Índia, Oman e Áustria	X				X		X		X

Figura 7. Métodos de visualização de dados utilizados pelos principais artigos revisados, que possuem resultados com foco na análise de projetos de EOR. Para cada artigo são listadas as revistas e ano em que foram publicados, números de citações obtidos no Google Scholar, com ordenação decrescente pelo número médio de citações por ano. Além disso, também são listadas as quantidades e abrangência das bases de dados utilizadas pelos artigos e dos países que contemplam.

4. METODOLOGIA

A sequência desenvolvida neste trabalho foi subdividida em cinco etapas principais, a saber: captura de informação e dados, pré-processamento e obtenção da base de dados consolidada, processamento, e respectivas aplicações para análises de interrelações entre as variáveis, desenvolvimento de obtenção de campos análogos, além de avaliações com base na literatura. Uma síntese da rotina metodológica desenvolvida está apresentada na Figura 8. Buscou-se nesse processo seguir um fluxo iterativo com melhores práticas sugeridas por metodologias como a KDD, a proposta por (HUBER et al., 2019) e a CRISP-DM, preconizada pela IBM (GREGORY PIATETSKY, 2014, 2014; SHAFIQUE; QAISER, 2014; WIRTH; HIPPE, 2000), de forma a garantir que os dados usados sejam relevantes e adequadamente processados para abordar as questões propostas.

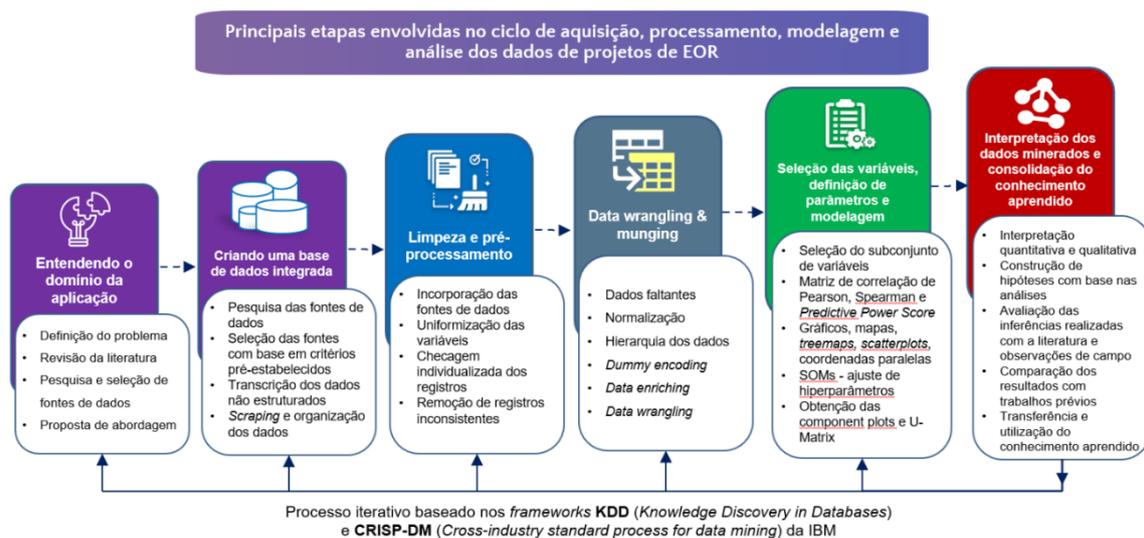


Figura 8. Síntese da rotina metodológica iterativa de mineração de dados, baseada nos modelos KDD e CRISP-DM, utilizada no presente trabalho.

4.1. Estruturação da base de dados

4.1.1. Origem e captura de informação

As principais fontes utilizadas podem ser subdivididas em:

- i. Artigos científicos considerados referências na aplicação de EOR, com maior número de citações no Google Scholar nesse contexto. Determinados artigos explicitam o estado da arte na aplicação de determinadas classes de EOR; outros, por sua vez, focam-se em reportar experiências de campos de produção em que foram implementados métodos de EOR
- ii. Levantamentos abrangentes e estudos da aplicação de métodos de EOR em reservatórios de determinadas regiões específicas do mundo ou determinadas bacias sedimentares, como do Mar do Norte (AWAN, 2008) de shale oil (ALFARGE, 2018; ALHARTHY et al., 2018; JIA; SHENG, 2017; WANG, 2017), projetos a partir de 2014 (LIU et al., 2020), carbonatos (ALVARADO; MANRIQUE, 2010), reservatórios *offshore* (BONDOR, 2005) e campos no Brasil, a exemplo de Peregrino na Bacia de Campos (HAN et al., 2012; KRISTOFFERSEN; STANKO; HOFFMANN, 2017), dentre outros.
- iii. Conjunto de experiências de campo acerca de um método em específico, como injeção de CO₂ associada com captura e armazenamento de carbono (LAKE; LOTFOLLAHI; BRYANT, 2019; OLEA, 2017; OUGHTON et al., 2016; SAINI, 2017), injeção alternada de água com gás (AWAN, 2008; BELAZREG, 2019; GBADAMOSI et al., 2018), uso de métodos de EOR microbiais (SHE et al., 2019), injeção de água de baixa salinidade (ARIANI, 2018; CHAVAN et al., 2019), injeção de produtos químicos (ALSOFI et al., 2019; MAHRAN; ATTIA; SAHA, 2018), combustão in-situ (TURTA et al., 2007), dentre outros.
- iv. Artigos com informações amplas, porém menos estruturadas ou com dados pontuais, a exemplo de (AISofI et al., 2019; Brodie et al., 2012; Correia and Schiozer, 2018; Jin, 2017; Meneses, 2018; Mitra et al., 2005; Morel et al., 2015; Pizarro and Branco, 2012; Seccombe et al., 2010; Sun et al., 2020, 2011);
- v. Relatórios e bases de dados de órgãos e institutos internacionais, como International Energy Agency (IEA), Center for Strategic and International Studies (CSIS), Global Carbon Capture and Storage Institute (CCSI), OPEC (Organization of the Petroleum Exporting Countries), dentre outros;
- vi. Órgãos reguladores e de pesquisa de diversos países;

- vii. Apresentações para investidores e relatórios técnicos de companhias operadoras dos campos; revistas e sites de notícias especializados no setor de petróleo e negócios;
- viii. Livros contendo estudos de casos e exemplos reais de aplicação de EOR (SHENG, 2013), (SHENG, 2010), e MANRIQUE (2014).

A utilização de várias fontes de dados neste trabalho ocorreu, sobretudo, pelos motivos elencados a seguir: (i) na abordagem orientada por dados há a inserção informação proveniente de diversas fontes e de projetos de diferentes localidades, em que a diversificação de fontes atenua possíveis vieses; (ii) a presença de grande quantidade de variáveis causa uma grande esparsidade de dados no espaço multidimensional, podendo levar ao efeito conhecido como mal da dimensionalidade (curse of dimensionality) (AGGARWAL, 2015; BERTINI; TATU; KEIM, 2011; GANGULI; SOMPOLINSKY, 2012) – e nesse caso a obtenção de mais dados reduz a porcentagem de dados faltantes; (iii) as fontes de dados de maneira isoladas não possuem todas as informações sobre os respectivos projetos de EOR que abordam, porém tendem a ser complementares; (iv) ao se utilizar de diferentes fontes, é possível a ampliar a representatividade e abrangência da base de dados frente ao universo total de projetos de EOR existentes no mundo.

4.1.2. Consolidação das fontes de dados

A base de dados Structured Query Language (SQL) compilada e processada contém 1237 instâncias, cada uma referindo-se a um projeto de EOR ao redor do mundo, contando com 323 variáveis passíveis de análise. Na estruturação da base de dados cada instância do conjunto de dados representou a execução de um projeto de EOR em um reservatório de petróleo no mundo, tendo como variáveis cinco grandes conjuntos de características: (i) registros de localização, (ii) litológicas, (iii) petrofísicas, (iv) características de projeto, constituindo dados reais da implementação de EOR no mundo, trazendo embutido consigo informações relevantes acerca dos desafios, dificuldades técnicas e demais fatores que podem influenciar.

4.1.3. Pré-processamento

Para tornar a base de dados o mais consistente possível, foi realizado o pré-processamento dos dados (Tabela 2), envolvendo uma sequência de etapas. Inicialmente foi feita a remoção de ruídos, incluindo informações espúrias e/ou inválidas e a uniformização das variáveis, conversões de unidades, tendo em vista em reuni-las em uma identidade numérica comum (de pés para metros, km² para acres, etc). Reservatórios duplicados foram mesclados ou removidos. Em uma inspeção geral do conjunto de dados, verificou-se possíveis inconsistências ou discrepâncias, bem como verificação de valores atípicos. Finalmente foram removidos dados que apontavam inconsistências, com checagens desenvolvidas a partir de pesquisas mais detalhadas e individualizadas sobre os respectivos campos. As variáveis obtidas e a estrutura geral da base de dados processada podem ser observadas na Tabela 2.

Tabela 2. Síntese da estrutura geral da base de dados e variáveis coletadas acerca de reservatórios de petróleo em que se teve aplicação de projetos de EOR.

Conjuntos de variáveis presentes na base de dados		
Registros de localização	Nome do campo	Anos de produção
	Nome da formação geológica	Mecanismo produção primária
	País	Tempo dos campos para chegar no platô de produção (anos)
	Estado/província	Tempo dos campos em produção no platô (anos)
	Região administrativa	Fator de recuperação atual (%)
	Condado/município	Watercut (%)
	Latitude e longitude (°)	EUR (%)
Propriedades básicas do reservatório	Litologia dominante	Escala do projeto
	Porosidade (%)	Data inicial e final do projeto
	Permeabilidade (mD)	Companhia operadora
	Profundidade (ft)	Ambiente <i>offshore</i>
	Temperatura (°F)	Lâmina d'água (ft)
	Net pay (ft)	Método de IOR/EOR:
Contexto geológico do reservatório	Tipo de bacia	- Tipo de fluido injetado
	Estilo estrutural	- Concentração e/ou PV inj.
	Ambiente de sedimentação	- Tipo de processo - Miscibilidade
	Idade geológica	- Associado a Carbon Capture & Storage
Propriedades do óleo	Tipo de selo	Nº de poços produtores
	°API	Nº de poços injetores
	Viscosidade (cP)	Padrão de injeção:

	Saturação inicial e final de óleo (%)		- Line, 4, 5, 7 or 9 spot
	Saturação final de óleo (%)		- Invertido, linedrive ou normal
	Pressão inicial (bars)		- Espaçamento entre os poços
	MMP (bars)		Densidade de poços (poços/acre)
	Número ácido (mg KOH)		Razão Nprod/Ninj
	Dykstra Parsons (0 a 1)		Eficiência de deslocamento
	TDS (ppm)		Avaliação do desempenho
Petrofísica especial	Soi (%) e Swi (%)	Eficiência da recuperação	Aumento da produção (bpd)
	Tipo e teor de argila		Fator de recuperação $1^{\circ}+2^{\circ}$
	Molhabilidade		Fator de recuperação IOR/EOR

Para tornar a estrutura das variáveis apropriada à aplicação das técnicas de IA foi feito o gerenciamento das variáveis (*Data Handling*) por meio de operações diversas a partir dos dados. O gerenciamento das variáveis, assim, incluiu a transformação de variáveis categóricas em binárias; a criação de novas variáveis com atributos ou grandezas físicas de interesse, a partir da combinação de propriedades anteriores, favorecendo novos parâmetros e enriquecendo a análise; a aplicação de transformação logarítmica para variáveis que apresentaram distribuição log-normal e normalização de variáveis, de forma a evitar diferenças nos pesos de treinamento das RNA's.

4.2. Processamento e metodologia de análise

4.2.1. Análises Multidimensionais por Self-Organizing Maps

Para a análise da interdependência entre as variáveis, a base de dados integrada foi inserida no algoritmo SOM, onde foi desenvolvido um espaço de dados de forma aleatória no conjunto n-dimensional. Uma grade hexagonal foi escolhida como formato de visualização. A superfície de um volume toroidal foi utilizada para a projeção dos neurônios ou BMU's para, respectivamente, manter relações de similaridade e topológicas, e evitar efeitos indesejáveis de borda. Foi atribuído um tamanho de mapa de 20 x 20 células. A inicialização foi definida como aleatória e os ajustes iniciais do treinamento conforme a Tabela 3. A Figura 9 mostra a sistemática aplicada às análises SOM.

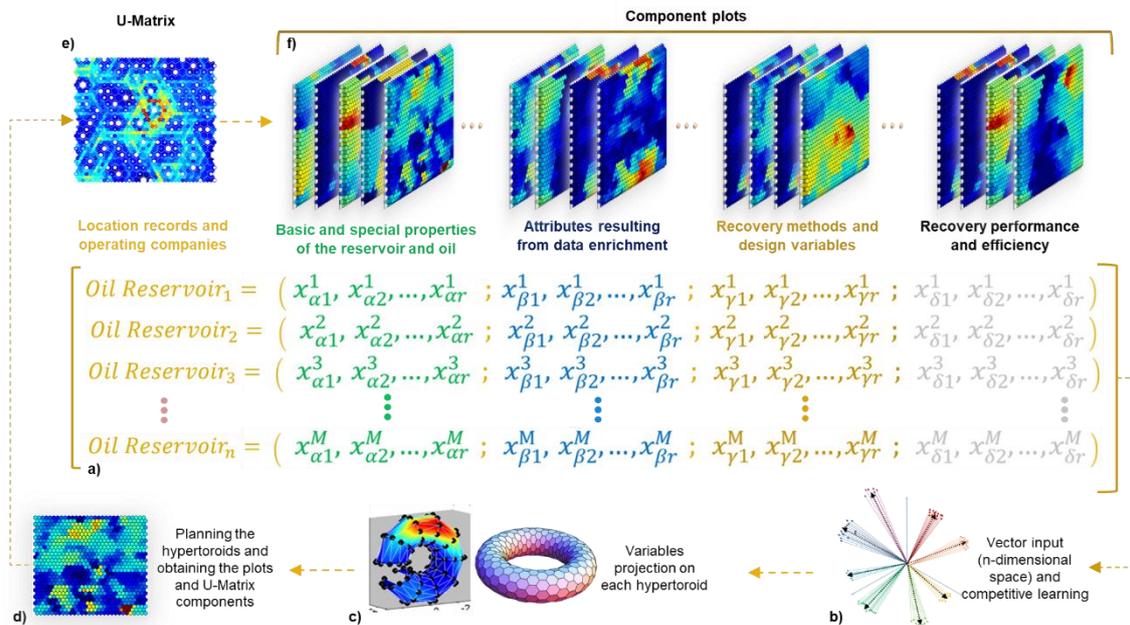


Figura 9. Síntese da aplicação do algoritmo SOM na base de dados estruturada obtida no presente trabalho, constituído das etapas: (a) base de dados processada, com as variáveis selecionadas e modeladas, é introduzida no algoritmo SOM; (b) o algoritmo SOM é inicializado, ocorrendo o processo de aprendizado e ajuste dos BMUs no espaço n-dimensional; (c e d) a rede neural tem os seus pesos projetados no hipertoróide e a U-Matrix (e) as *component plots* (f) são obtidas a partir da planificação dos respectivos hipertoróides. A U-Matrix representa as dissimilaridades no ambiente SOM dos BMUs e as *component plots* a influência de cada variável para cada conjunto de amostras representadas por um BMU no espaço SOM.

Após a geração do mapa auto-organizado, são produzidas as imagens 2D representativas do espaço n-dimensional dadas pelas *component plots* e pela *U-Matrix*. As *component plots* possibilitam visualizar e quantificar a contribuição das variáveis analisadas para cada neurônio resultante no mapa auto-organizável, sendo possível verificar interrelações entre as respostas das variáveis analisadas. A matriz-U possibilita, então, a classificação dos dados relacionada ao vetor similaridade, construído a partir das instâncias originais.

Cada neurônio que possua instâncias de reservatórios associados no mapa auto-organizado foi representado por um BMU, ou seja, um “mini-cluster” das instâncias de entrada, agrupando reservatórios ou campos de produção com características análogas. Como resultados, teremos os BMU's para cada reservatório analisado.

Tabela 3. Hiperparâmetros de treinamento do algoritmo SOM.

Ajustes iniciais de treinamento do algoritmo			
Grosso		Fino	
Raio Inicial	20	Raio Inicial	17
Raio Final	9	Raio Final	1
Extensão do Treinamento	20	Extensão do Treinamento	40 0

Como parte da análise dos resultados gerados, foi realizado um procedimento integrado, fazendo uso tanto de interpretações quantitativa quanto qualitativas. São utilizados de maneira conjunta scatterplots, histogramas e correlação entre variáveis junto à observação das component plots geradas pelo SOMs. Hipóteses são construídas com base na análise conjunta das variáveis, buscando avaliação das inferências realizadas com a literatura e observações de campo. Tal processo é realizado de maneira cíclica, conforme preconizado pelas metodologias CRISP-DM e KDD, com ampliação contínua da base de dados, reprocessamento e novas modelagens e análises dos dados.

4.2.2. Obtenção de campos análogos

A partir da base de dados estruturada e pré-processada a sequência desenvolvida para a obtenção dos campos análogos consistiu de quatro etapas principais: (i) seleção de parâmetros prioritários; (ii) processamento, com aplicação de algoritmo de ranqueamento de similaridade e de clusterização hierárquica e obtenção de conjunto de campos análogos; (iii) avaliação, com validação estatística do conjunto de análogos obtido com reservatório de interesse, e (iv) análises comparativas e qualitativas e detalhadas de análogos obtidos.

Uma síntese da rotina metodológica desenvolvida está ilustrada na Figura 10.

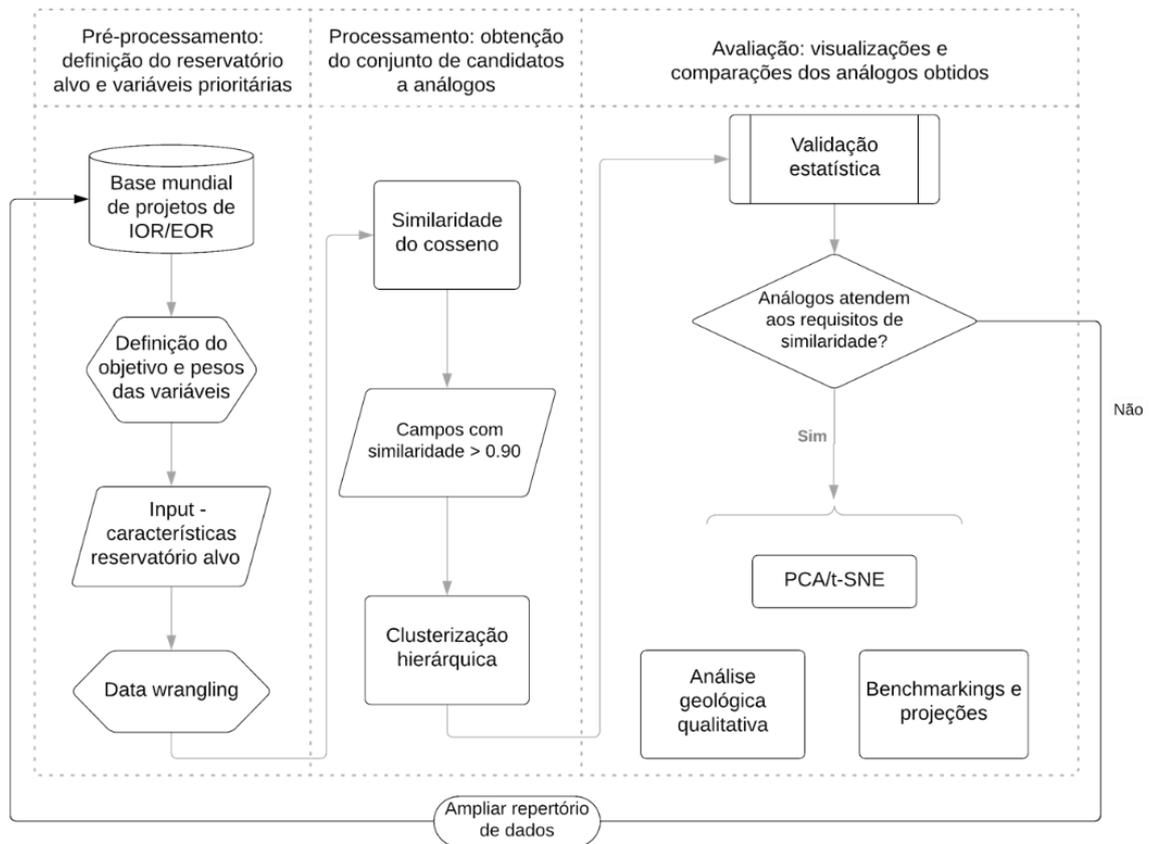


Figura 10. Rotina metodológica desenvolvida para a obtenção e análise de campos análogos.

4.2.2.1. Seleção de parâmetros prioritários

A definição dos parâmetros prioritários varia conforme o contexto de aplicação. Para o exemplo aqui desenvolvido, escolheu-se o campo de Serra, localizado na Bacia do Potiguar, no nordeste brasileiro. É um campo classificado maduro, com cenário de desenvolvimento similar a outros campos que vêm sendo adquiridos por operadoras brasileiras de médio e pequeno porte para sua revitalização. Para tanto, a ANP demanda a apresentação de campos análogos tanto para submissão de planos de desenvolvimento como para o pedido de redução de royalties por parte das operadoras (resoluções da ANP N°47 de 2014 e N°749 de 2018, e a instrução normativa N°2 de 2020), e a seleção de parâmetros nesse exemplo para obtenção de análogos busca contemplar os requisitos estabelecidos pela agência em que define reservatório análogo como “reservatório com propriedades de rocha e fluidos, condições de Reservatório (profundidade, temperatura e pressão) e mecanismos de produção similares, porém, geralmente, em estágio mais avançado de desenvolvimento do que do

Reservatório de interesse, podendo, desta forma, fornecer conceitos para auxiliar na interpretação de dados e na estimativa de recuperação” (resolução ANP nº47 de 2014).

O campo de Serra consiste de reservatórios areníticos situados na Formação Açú. As características do campo e os atributos definidos como prioritários para a busca de análogos encontram-se nas Tabelas 2 e 3.

Tabela 4. Variáveis categóricas consideradas no processo de busca de campos análogos para o reservatório de interesse. Marcadas com * e em negrito encontram-se as características do campo de Serra e que são desejáveis de se encontrar nos campos análogos a serem obtidos.

Litologia dominante	Mecanismo produção primária	Ambiente de sedimentação	Estilo estrutural	Tipo de selo	Tipo de bacia	Tipo de trapa	Idade geológica	Anos de produção
*Arenito	Gás em solução	Turbidito	Compressional	Sal	Rift	*Estrutural	Neógeno	Até 10
Arenito não consolidado	Capa de gás	Delta	*Distensional	*Folhelho	*Margem passiva	Estratigráfica	Paleógeno	10 a 20
Dolomita	*Influxo de água	Fluvial	Transpressional	Mudstone	Intra-cratônica	Mista	*Cretáceo	20 a 30
Calcário	Descompressão	Leques aluviais	Transtensional	Ígnea	Strike-slip	Domo de sal	Jurássico	*30 ou mais
Conglomerado		Tidal			Fore arc	Anticlinal	Triássico	
		Eólico			Back arc		Permiano	
		Glacial			Foreland		Carbonífero	
		Vulcânico			Sag		Devoniano	
		Praia					Siluriano	
		*Fluvial/ delta/estuário					Ordoviciano	
		Fluvial/leques aluviais						

Tabela 5. Variáveis numéricas de características dos reservatórios do campo de Serra e que foram consideradas no processo de busca de campos análogos.

Reservatório Variável	Serra (óleo pesado)	Serra (óleo leve)
Litologia dominante	Arenito consolidado	Arenito consolidado
Porosidade média (%)	24	22
Permeabilidade média (mD)	676	1062
Profundidade (m)	810	877
Temperatura (°C)	51	52.5
Net pay (ft)	20	22.5
Grau API	17.3	30.5
Viscosidade (cP)	40	7.5
Saturação do óleo (%)	51	44

A obtenção de análogos e construção dessas análises se dá por meio de duas principais etapas, consistindo em um *drilldown* com a realização de análises mais detalhadas sobre os campos candidatos a análogos à medida em que se avança na metodologia. O campo de Serra possui dois compartimentos principais de produção: um com óleo pesado (°API aproximadamente 17) e outro com óleo leve (°API aproximadamente 31). Nesse sentido, serão buscados análogos para cada um desses casos considerando as variáveis elencadas acima.

4.2.2.2. Ordenação por similaridade do cosseno e clusterização hierárquica

A busca por campos análogos é uma análise multicritério que deve incluir uma série de variáveis, englobando propriedades geológicas, petrofísicas, de fluido, engenharia e de projeto. Quanto maior o grau de similaridade dos campos candidatos a análogos com o reservatório de interesse, mais acuradas serão as projeções e insights que o campo análogo poderá fornecer. São considerados mais de 30 atributos, englobando tanto variáveis categóricas quanto variáveis numéricas. Na busca de candidatos a análogos duas técnicas complementares são utilizadas a partir da base de dados mundial: ranqueamento via similaridade do cosseno e clusterização hierárquica.

A similaridade do cosseno mede a semelhança entre dois vetores por meio do produto escalar interno normalizado do espaço (Hand, 2014), conforme a equação (1):

$$\begin{aligned} \text{similaridade do cosseno} = S_c(A, B) &:= \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \end{aligned}$$

Pode ser entendida como o cosseno do ângulo no plano formado entre dois vetores, permitindo determinar o quão dois vetores estão apontando na mesma direção, conforme ilustra a Figura 11. É uma medida que permite considerar, ao mesmo tempo, tanto variáveis numéricas quanto categóricas no mesmo modelo, o que se faz necessário no contexto das propriedades dos reservatórios, que para o objetivo proposto envolvem tanto atributos numéricos (porosidade, permeabilidade, °API, etc) quanto categóricos (litologia dominante, ambiente de sedimentação, tipo de bacia, etc). Dessa forma, a partir do cálculo da similaridade do cosseno entre os pares de vetores que representam o reservatório de interesse e campos candidatos a análogos, os reservatórios são ordenados conforme maior grau de similaridade em relação ao reservatório de interesse do estudo de caso e então são selecionados os 50 com maior semelhança para análises comparativas.

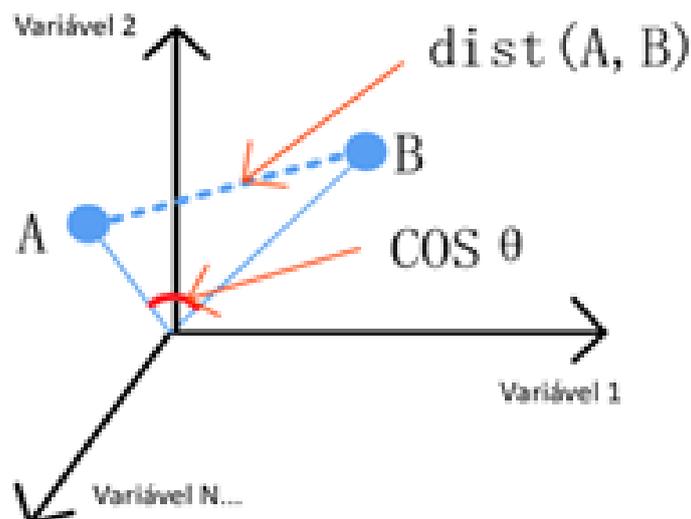


Figura 11. Representação das medidas de similaridade distância euclidiana e ângulo do cosseno para dois pontos A e B.

A clusterização hierárquica consiste em algoritmos que agrupam conjuntos de dados semelhantes baseado em medidas (?) de XYZ. O endpoint

é um conjunto de *clusters*, onde cada cluster é distinto um do outro cluster e os objetos dentro de cada cluster são amplamente semelhantes entre si (Murtagh & Contreras, 2012). Como ilustrado na Figura 12, uma forma conveniente de representar tal estrutura interna dos dados é via dendrograma, em que cada nó na árvore do cluster contém um grupo de dados semelhantes; os nós se agrupam conforme o grau de similaridade com os outros nós, e os *clusters* em um nível se unem a *clusters* próximos no nível acima; o processo continua até que todos os nós estejam na árvore. Espera-se que os campos de petróleo dentro do mesmo cluster compartilhem características geológicas e de comportamento de fluido semelhantes, permitindo traçar análises complementares de similaridade à provida pela similaridade do cosseno.

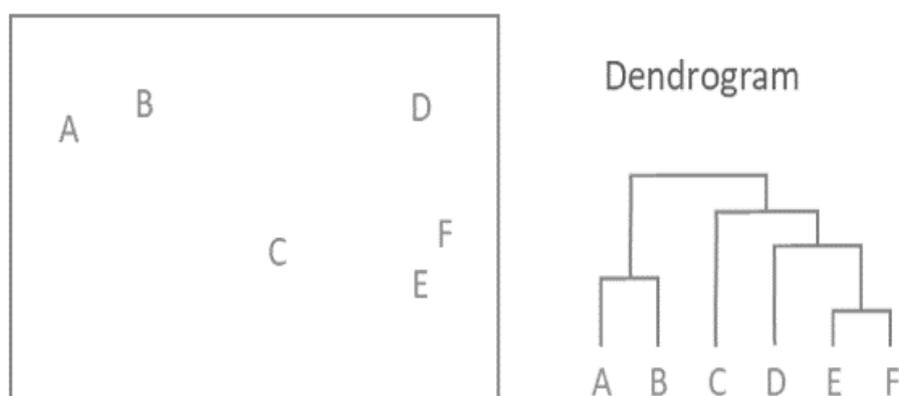


Figura 12. Ilustração de agrupamento realizado por clusterização hierárquica e sua representação na forma de dendrogramas a partir de amostras dispostas no espaço bidimensional.

4.2.2.3. Avaliação e comparação dos análogos obtidos

Como método de avaliação são utilizados parâmetros estatísticos, como o desvio em relação ao histograma, e análises comparativas e qualitativas que permitem indicar o quão similar são os candidatos a análogos obtidos em comparação com o caso tomado como base, tendo como contexto o conjunto mundial de reservatórios, e tomando como base as variáveis definidas como de maior prioridade. A avaliação utiliza como base critérios e variáveis estabelecidas por diretrizes internacionais, como o PRMS, e por agências reguladoras como a ANP para a apresentação de campos análogos exigidos na submissão de planos de desenvolvimento e de redução de royalties por parte das operadoras (resoluções da ANP N°47 de 2014, N°749 de 2018 e instrução

normativa N°2 de 2020). Caso haja um desvio das propriedades obtidas em relação maior do que o tolerado, realiza-se um novo ciclo de aquisição e processamento de dados de campos que possuam potencial de apresentarem maior similaridade com o reservatório alvo. Esse processo é essencial para o controle de qualidade dos resultados obtidos, com a obtenção de análogos validados em relação a níveis de similaridade com o reservatório de interesse e para possibilitar extrapolações e análises comparativas avançadas.

A etapa inicial fornece uma lista e grupos de campos com maiores similaridades ao do reservatório de interesse. São realizadas análises de forma a explorar os padrões internos do conjunto de análogos obtido. É executado algoritmo de redução de dimensionalidade PCA, que permite a visualização em um espaço de baixa dimensionalidade do conjunto de reservatórios análogos. A aplicação das técnicas de clusterização hierárquica, com a obtenção de dendrogramas, também possibilita a identificação inteligente de tendências de agrupamento entre grupos de reservatórios similares.

São realizadas na sequência análises geológicas detalhadas e qualitativas de quatro campos análogos tomados como exemplo. São analisadas as semelhanças e diferenças desses campos com o reservatório alvo. Nessa análise são comparadas de forma mais aprofundada parâmetros relativos ao contexto geológico, de desenvolvimento e produção desses campos utilizados como exemplo, considerando também outros atributos coletados, mas que não foram incluídos no processo de busca semi-automatizado. Também são desenvolvidos benchmarkings e análises comparativas, além da identificação de estratégias de recuperação empregadas em campos similares, que possibilitem repensar o desenvolvimento de um campo de petróleo de forma a se otimizar a exploração e o aumento do fator de recuperação.

5. RESULTADOS

5.1. Visão geral da base de dados

A base de dados integrada resultante conseguiu integrar 323 variáveis relativas aos 1237 reservatórios, operados por 96 empresas, distribuídos em 36 países. Conforme mostra a Figura 13, os projetos coletados englobam a aplicação de 4 grandes classes e 24 subclasses de EOR.

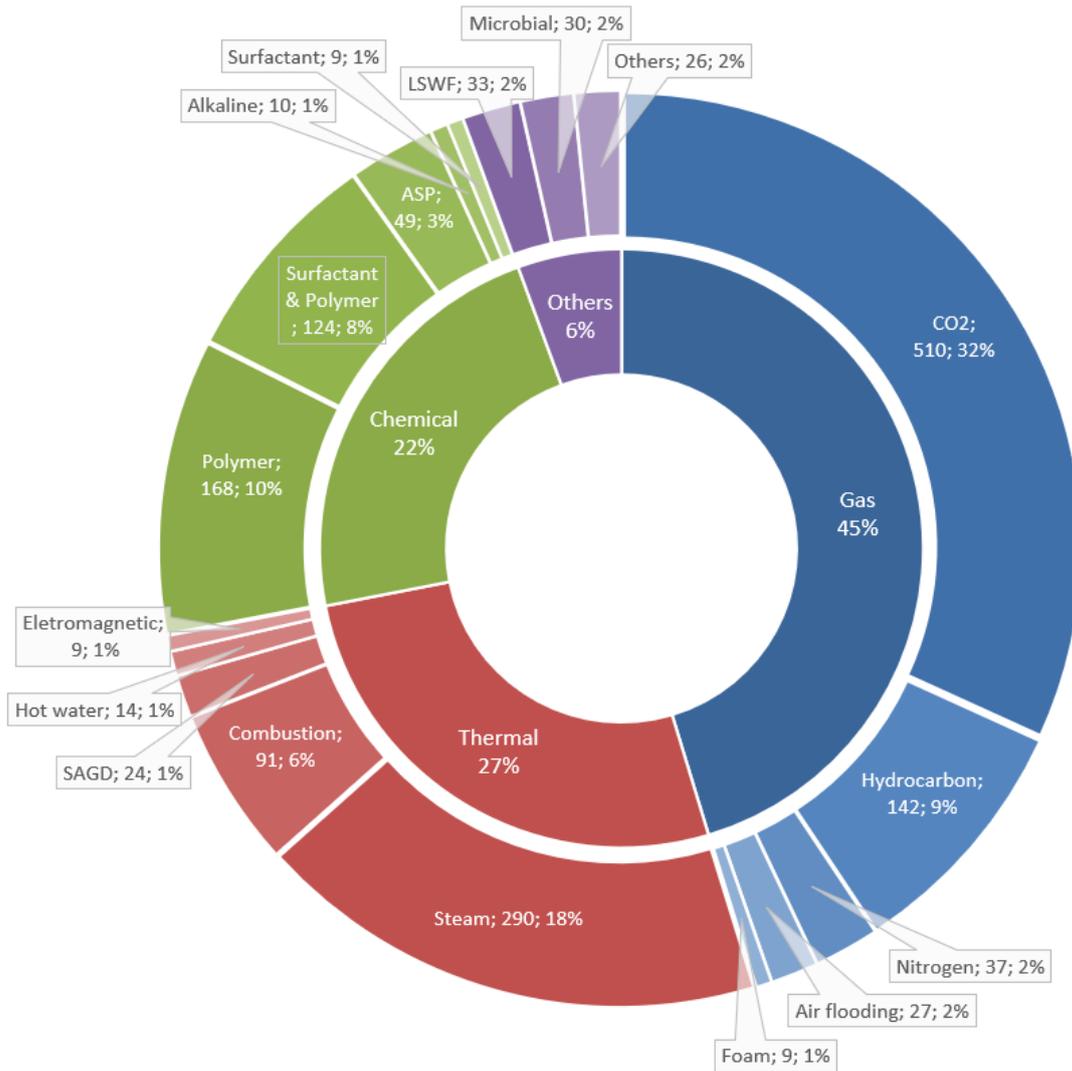


Figura 13. Constituição da base de dados em relação às classes de métodos de EOR presentes.

5.1.1. Localização e histórico dos projetos de EOR

Observa-se a existência de projetos de aplicação de EOR com significativa distribuição global, revelando a disseminação das técnicas de EOR. Porém, como ilustram as Figuras 14 e 16, nota-se uma concentração das aplicações destas técnicas em determinadas regiões. Dentre tais, destacam-se os EUA,

Canadá, China, países banhados pelo Mar do Norte, Venezuela, Brasil, dentre outros. As regiões de maior ocorrência se distribuem desde a América do Norte e Europa, com grande demanda energética e com um já grande número de projetos de EOR implantados há décadas, passando pela China, com intensa aplicação recente, até países da América Latina, Oriente Médio e África com uma escala relativamente menor de adoção de técnicas avançadas de recuperação de petróleo.

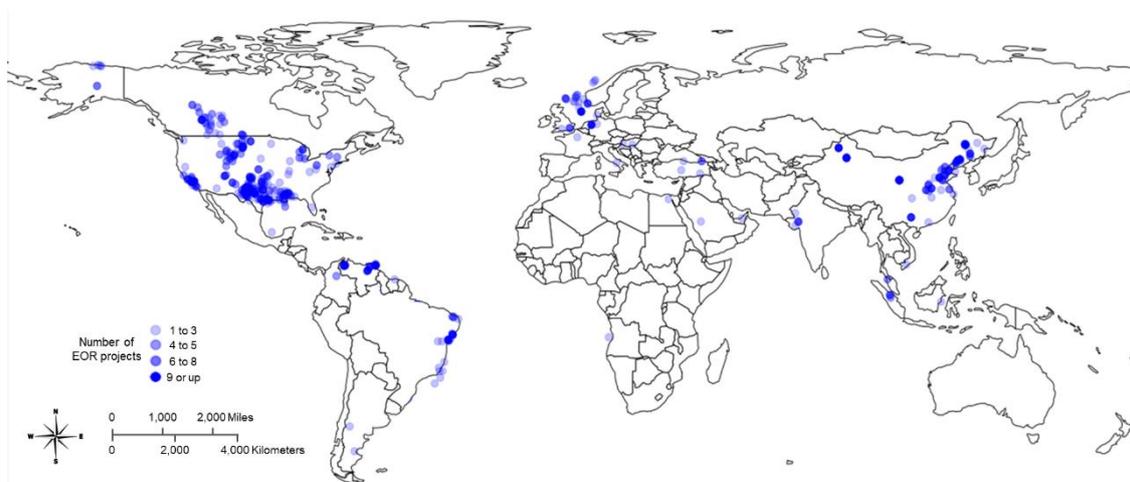


Figura 14. Localização dos campos de petróleo que tiveram aplicação de EOR presentes na base de dados aqui apresentada.

A Figura 15 traz a frequência dos projetos para cada classe de método de EOR ao longo dos anos no intervalo entre 1959 e 2015. À distribuição foi sobreposta a série temporal de preços do barril de petróleo nominal e descontado a inflação tendo como referência os EUA. O preço descontado a inflação também fora incluído para se obter uma comparação mais justa entre cenários econômicos distintos. Nota-se que a frequência de início de projetos de EOR de forma geral acompanha a curva de preço do petróleo. Porém, entre as décadas de 1980 e 1990, a frequência manteve-se elevada apesar da queda no preço do petróleo. Essa variação ocorreu, provavelmente, por políticas de incentivo do Departamento de Energia dos EUA para a implementação de projetos de EOR. Além disso, notou-se uma baixa frequência de projetos com data inicial acima de 2014. Contudo, não é possível afirmar que na última década houve poucos projetos de EOR implementados. A baixa frequência constatada pode

simplesmente refletir que os dados acerca da aplicação de EOR mais recentes ainda não foram divulgados pelas empresas operadoras ou publicados em artigos científicos. Dessa maneira, estes projetos mais recentes não foram contemplados na estruturação da base de dados no presente momento. Como não foi possível se obter as datas finais da implementação dos projetos de EOR para parte considerável dos campos, optou-se pela frequência de projetos de EOR por ano no gráfico levar em consideração apenas suas respectivas datas de início.

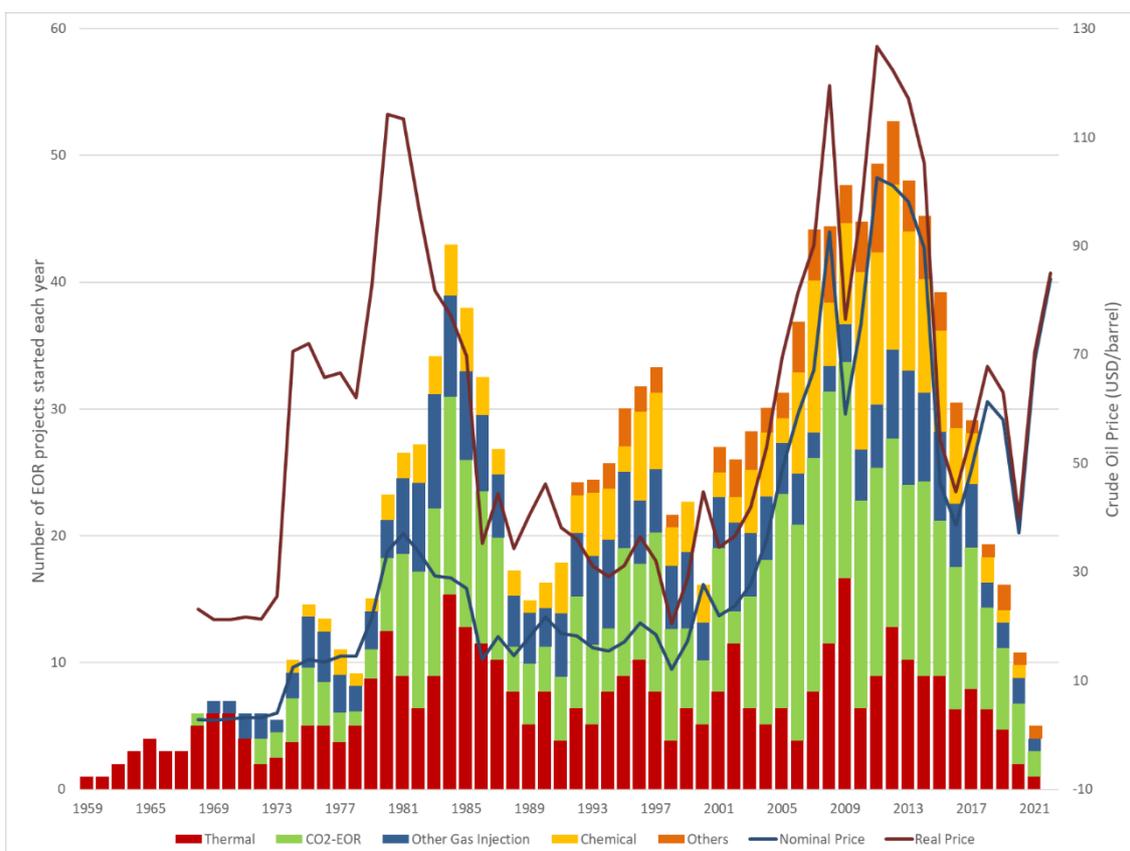


Figura 15. Barras de frequência de projetos para cada categoria de EOR ao longo dos anos sobrepostas pela curva relativa ao preço do barril de petróleo Brent, nominal e descontada a inflação, tendo como base índice econômico dos EUA.

Algumas classes de métodos de EOR possuem maior aplicação em determinados países, conforme pode ser observado na Figura 16. Em especial, nota-se que os EUA lideram a aplicação de injeção de vapor, seguido pela Venezuela, China, Canadá e outros países. O maior número de projetos e as maiores produções de petróleo decorrente da implementação de métodos de EOR ocorreram nos EUA. Dentre os motivos para isso, destaca-se o fato da

maturidade dos EUA na aplicação de métodos de EOR, com um grande número de campos de petróleo depletados e políticas de incentivo estaduais e federais. Além disso, o país apresenta alta disponibilidade de tecnologia e infraestrutura, tornando-o lugar ideal para implementação de projetos pilotos e comerciais de EOR, mesmo em períodos de baixos preços do barril de petróleo.

Nota-se em especial um aumento do número de publicações e de projetos de EOR relacionados à injeção de produtos químicos, em especial à injeção de polímero e surfactante. Enquanto nas décadas de 1980 e 1990 houve, no máximo, 5 projetos de EOR químicos por ano, nas décadas de 2000 e 2010 esse número chegou a superar 15 projetos; para a frequência de publicações ao longo dos anos, tal crescimento fica ainda mais evidente. Tal fenômeno pode ser explicado, especialmente na China, por incentivos regulatórios, barateamento de produtos de química fina e pelo desenvolvimento de novos compostos que motivaram o estudo e a aplicação em variadas condições de reservatório. Esse fato revela tanto o emprego apenas mais recentemente, a partir dos anos 2000, dos métodos químicos em maior escala, como também a liderança chinesa em tal classe de EOR. Para a injeção de gás, os EUA lidera o número de projetos executados, seguido pelos países do Mar do Norte.

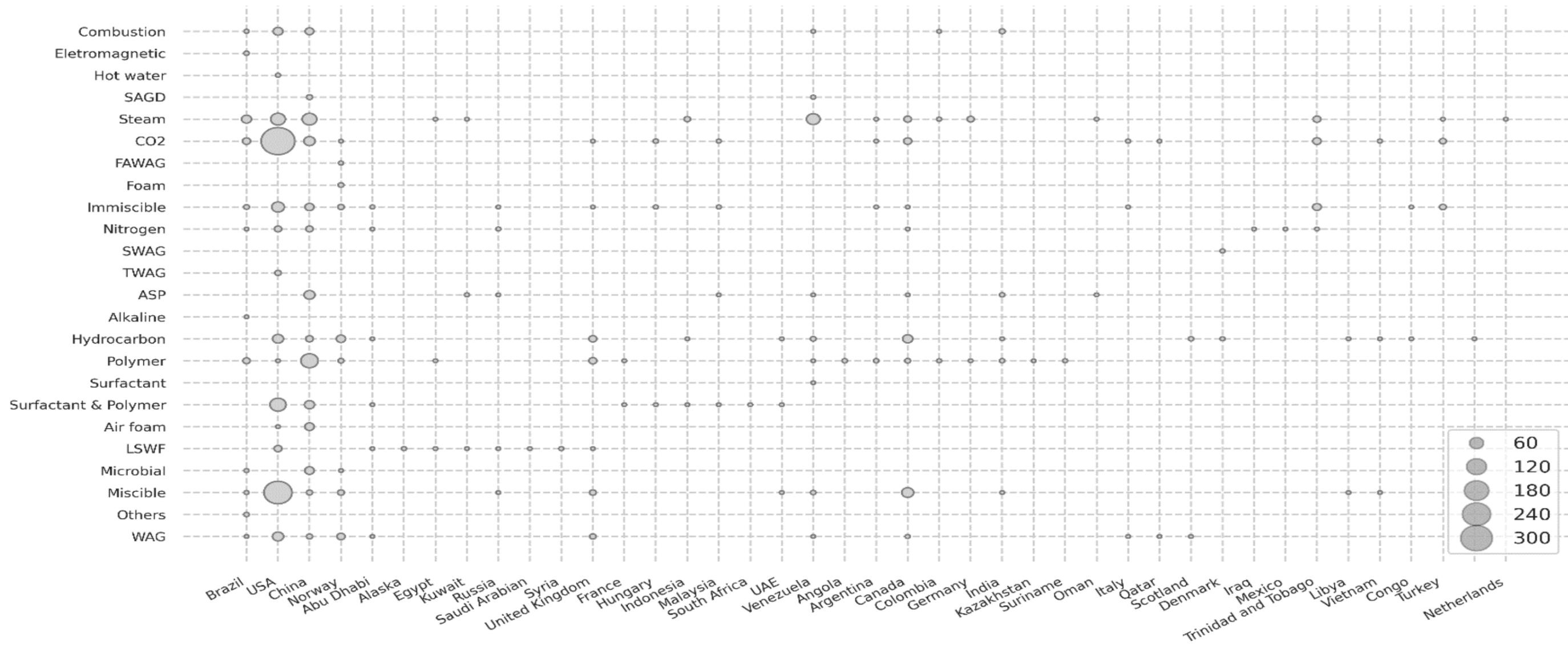


Figura 16. Categorical scatterplot com a distribuição das subclasses de métodos de EOR vs os países em que foram aplicados. O tamanho dos pontos é proporcional à frequência acumulada de projetos do método para com o respectivo país.

Os projetos de EOR coletados em escala mundial foram ou são conduzidos por diversas companhias operadoras atuantes da indústria de petróleo. Há empresas globalmente conhecidas e integradas ao longo de toda a cadeia produtiva de petróleo e de comercialização de derivados, a exemplo das operadoras *British Petroleum* (BP), *Chevron*, *China National Offshore Oil Corporation* (CNPC), *Equinor*, *ExxonMobil*, *Petrobras*, *Petronas*, etc. Há também empresas independentes e muitas vezes com atuação regional, tais como *Anadarko*, *Apache*, *Athabasca*, *Canadian Natural Resources* (CNRL), *Carrizo*, *Continental Resources*, etc, mas que representam por significativa parcela da liderança na execução de projetos de EOR.

A partir da *treemap* na Figura 17 é possível observar a tendência de métodos químicos terem sua aplicação realizada majoritariamente por companhias integradas, sejam *National Oil Companies* (NOCs) ou *International Oil Companies* (IOCs). Já os métodos de injeção de gás tiveram seus projetos executados não apenas por companhias integradas NOCs e IOCs, mas também por significativa proporção de empresas classificadas como médias e independentes no setor de E&P. Para os métodos térmicos também há uma diversidade de tipos de empresas atuando, englobando desde NOCs e, em menor proporção, IOCs, e até operadoras médias e incluindo pequenas empresas focadas em E&P.

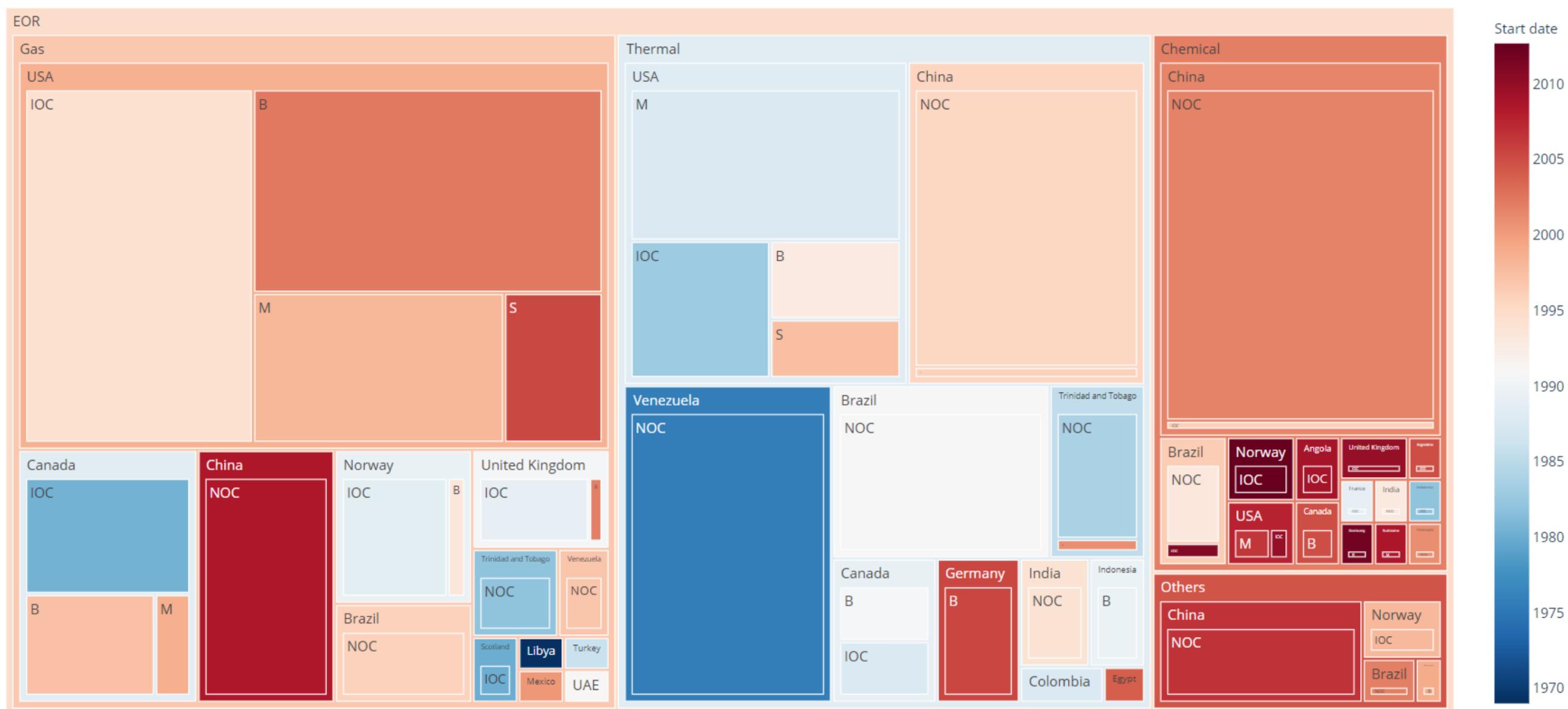


Figura 17. *Treemap* representando a distribuição dos projetos de EOR, agrupados conforme classe de EOR, o país de localização e classificação da empresa operadora do campo, conforme ela seja IOC (International Oil Company), Big (B - receita anual entre 5 e 20 bilhões de dólares), Medium (M - receita anual entre 500 milhões e 5 bilhões de dólares) e Small (S - receita anual abaixo de 500 milhões de dólares). A cor de cada retângulo refere-se à data média de início dos projetos, indicada pela escala ao lado.

5.1.2. Parâmetros de reservatórios

Com base no scatterplot da Figura 18, nota-se comportamentos distintos para as classes de EOR quanto à relação entre profundidade médias do reservatório e datas de início dos projetos. Primeiramente, pode ser notada uma correlação geral positiva de projetos mais recentes apresentarem maiores profundidades de reservatório. Em segundo lugar, observa-se tal correlação positiva para os métodos químicos e de injeção de gás. Já os projetos de métodos térmicos estes apresentam comportamento distinto, mostrando uma aparente barreira no avanço para profundidades maiores que 5000 ft quando comparados às demais classes de EOR.

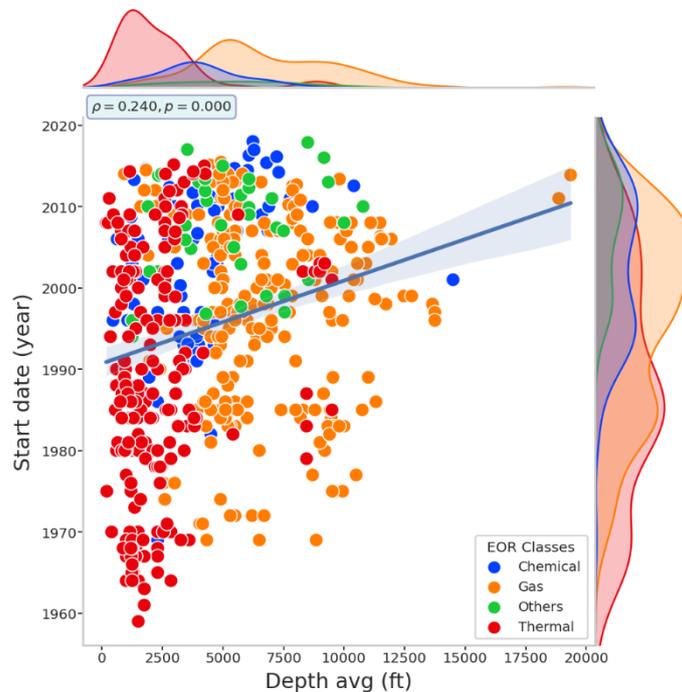


Figura 18. *Scatter plot* das variáveis profundidades dos reservatórios e datas de início dos projetos, com categorização dos dados quanto à classe do método de EOR utilizada.

A partir de parâmetros petrofísicos foi calculado o parâmetro sintético da razão logarítmica entre permeabilidade e porosidade, e que possui relação com tamanho estimado da garganta de poro e respectiva litologia (JERAULD; SALTER, 1990; KOLODZIE, 1980; REZAEI; JAFARI; KAZEMZADEH, 2006). Com base na Figura 19, nota-se que os métodos térmicos tendem a ser aplicados em reservatórios com garganta estimada de poro maior e óleos de índices de mobilidades menores e mais viscosos se comparados a métodos químicos e de injeção de gás. Os métodos de injeção de gás e produtos químicos

têm maior aplicação em óleos mais leves e com gargantas de poros menores e os métodos químicos em valores intermediários para essas variáveis.

Nota-se duas tendências de disposição dos dados, representadas pelas retas *r* e *s*. Há a tendência de que os reservatórios com injeção de gás se disponham ao longo da linha tracejada *s* em azul, com os reservatórios que tiveram métodos químicos pouco abaixo desta linha; já os métodos térmicos tendem a se dispor ao longo da linha *r* preta. Tal fato indica que para a injeção de gás quanto maior o tamanho estimado da garganta de poro, maior tende a ser a mobilidade do óleo desses reservatórios. Já para os métodos térmicos, há uma relação inversa: quanto maior a garganta de poro, a mobilidade do óleo desses reservatórios tende a ser menor. Por inspeção nota-se agrupamentos dos dados de acordo com a litologia dominante de cada reservatório ao se correlacionar as variáveis porosidade e tamanho estimado da garganta de poro. Em especial, arenitos tendem a apresentar maiores valores de porosidade e tamanho estimado da garganta de poro, enquanto carbonatos possuem uma maior dispersão dos dados, porém com concentração em valores menores de porosidade e tamanho estimado da garganta de poro. Nesse sentido, projetos de injeção de vapor tendem a serem aplicados em reservatórios em geral areníticos com elevados tamanhos estimados da garganta de poro e permeabilidades, mas que apresentam óleo de elevada viscosidade, o que acarreta num baixo índice de mobilidade conforme visualizado na Figura 19.

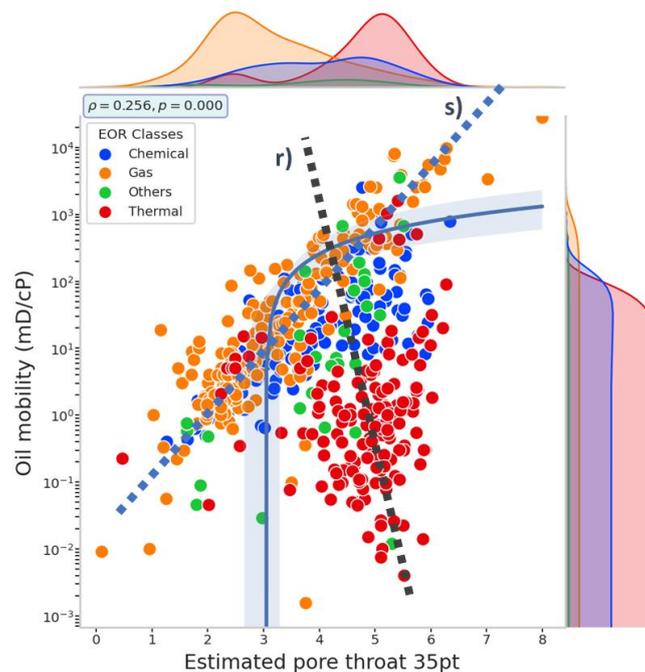


Figura 19. Joinplot com a plotagem cruzada das variáveis mobilidade estimada do óleo e tamanho estimado da garganta de poro. As retas r) e s) indicam tendências de correlação distintas para subgrupos de dados.

Os *boxplots* da Figura 20 resumam a distribuição de parte das variáveis técnicas em relação aos métodos de EOR pertencentes a cada classe principal. É possível observar que, para uma mesma classe de EOR, há variações significativas dos ranges de aplicação das variáveis técnicas conforme o método de EOR utilizado. Tais diferenciações entre os métodos de recuperação serão abordadas na seção “Análises por SOM dos métodos de EOR”.

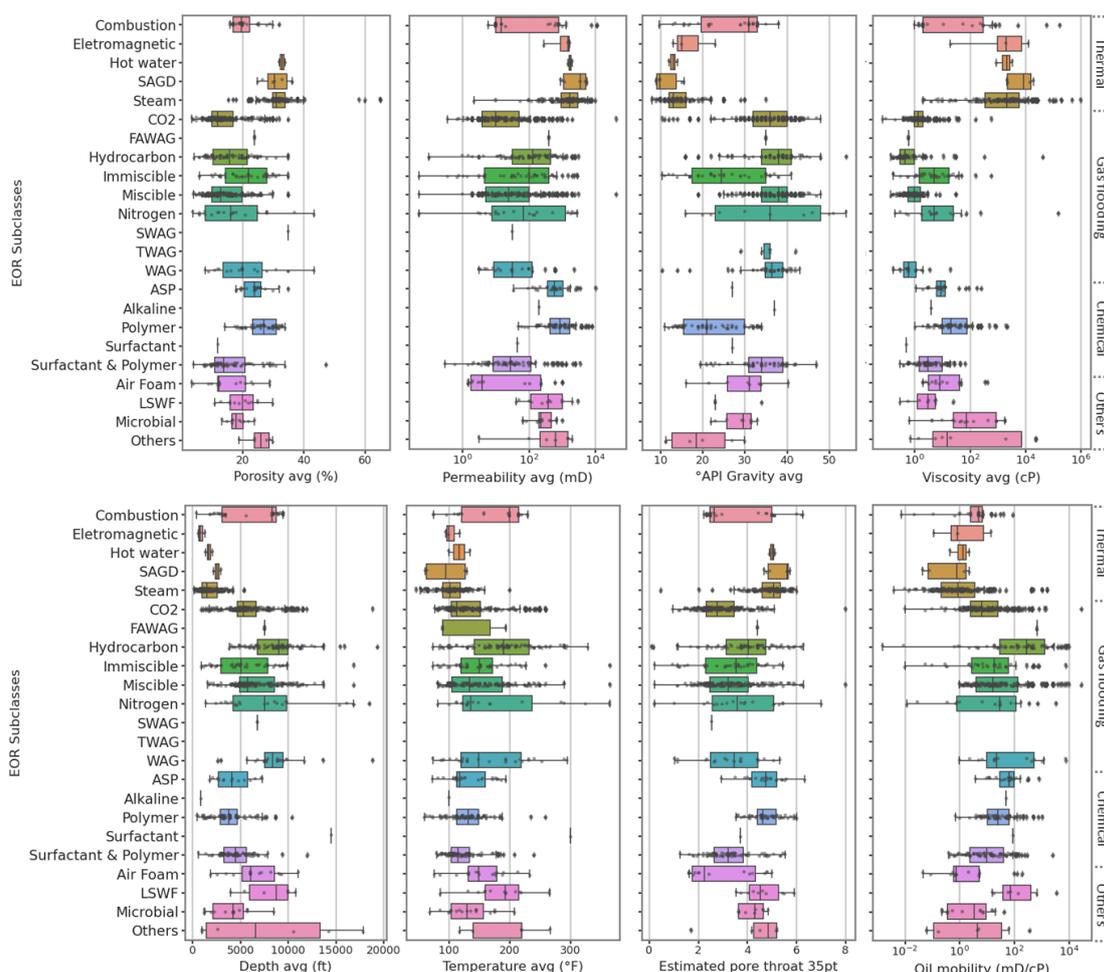


Figura 20. *Boxplots* para os principais tipos de métodos de EOR para diferentes variáveis de natureza técnica e petrofísica.

5.1.3. Variáveis de projeto

5.1.3.1. Companhias operadoras

Os projetos de EOR coletados em escala mundial foram ou são conduzidos por uma diversa gama de companhias operadoras atuantes da indústria de petróleo. Verifica-se no mundo grande diversidade de empresas na implementação de projetos de EOR, havendo desde empresas globalmente conhecidas e integradas ao longo de toda a cadeia produtiva de petróleo e de comercialização de derivados, mas que representam por significativa parcela da liderança na execução de projetos de EOR.

A partir da *treemap* na Figura 21 é possível observar a tendência de métodos químicos terem sua aplicação realizada majoritariamente por companhias integradas, NOC e IOC. Já os métodos de injeção de gás tiveram seus projetos executados tanto por companhias integradas NOC e IOC, mas também há uma porção significativa de empresas classificadas como médias e independentes no setor de E&P. Para os métodos térmicos também há uma diversidade de tipos de empresas atuando, englobando desde NOC e, em menor proporção, IOC, até operadoras médias e incluindo pequenas empresas focadas em E&P. Enquanto a Figura 21 faz o agrupamento das classes de companhias conforme o seu porte e atuação, as Figura 22 e 23 trazem especificado quais companhias atuam em quais países e com quais tipos de métodos de EOR, tendo como escala de cor a data de início dos projetos e a profundidade média dos reservatórios, respectivamente.

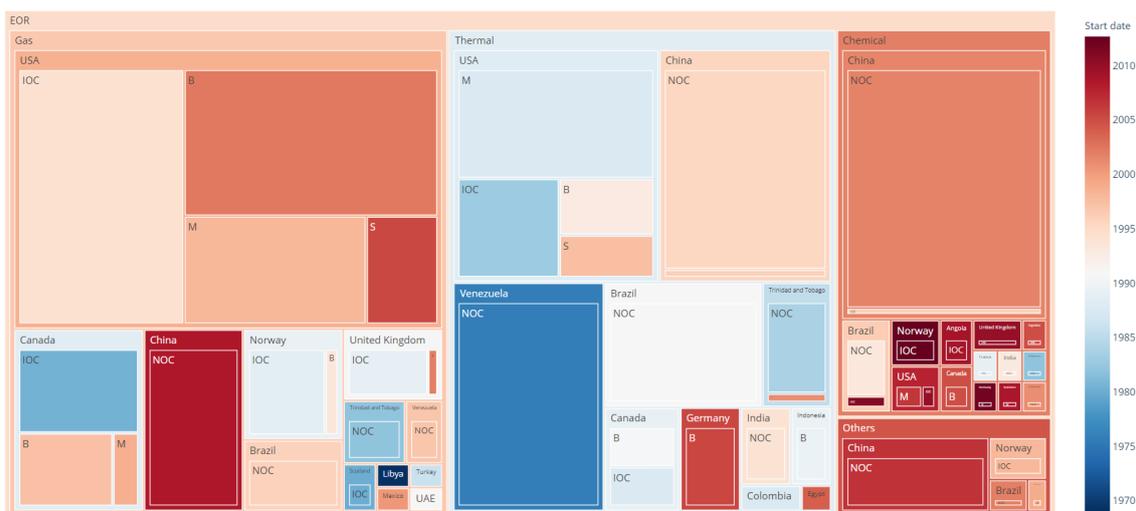


Figura 21. *Treemap* representando a distribuição dos projetos de EOR, agrupados conforme classe (injeção de gás, térmico, químico ou outros), o país de localização e classificação da empresa operadora do campo, conforme ela seja IOC (International Oil Company), B (Big, com receita anual entre 5 e 20 bilhões de dólares), M (Medium, com receita anual entre 500 milhões e 5 bilhões de dólares) e S (Small, com receita anual abaixo de 500 milhões de dólares). Além

disso, a cor de cada retângulo refere-se à data médio de início dos projetos, indicada pela escala ao lado.

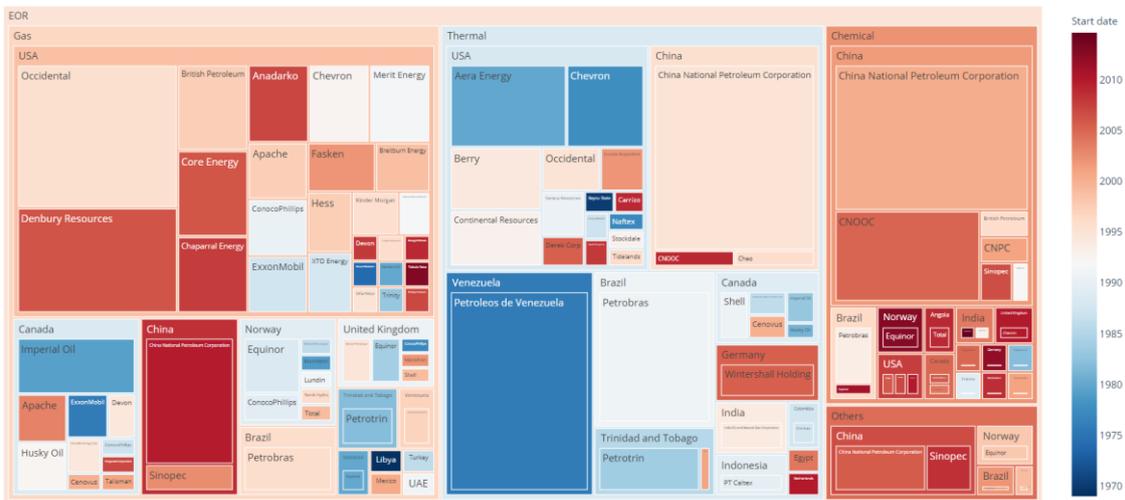


Figura 22. *Treemap* representando a distribuição dos projetos de EOR, agrupados conforme hierarquia de classe (injeção de gás, térmico, químico ou outros), do país de localização e da empresa operadora do campo. Além disso, a cor de cada retângulo refere-se à data médio de início dos projetos, indicada pela escala ao lado.

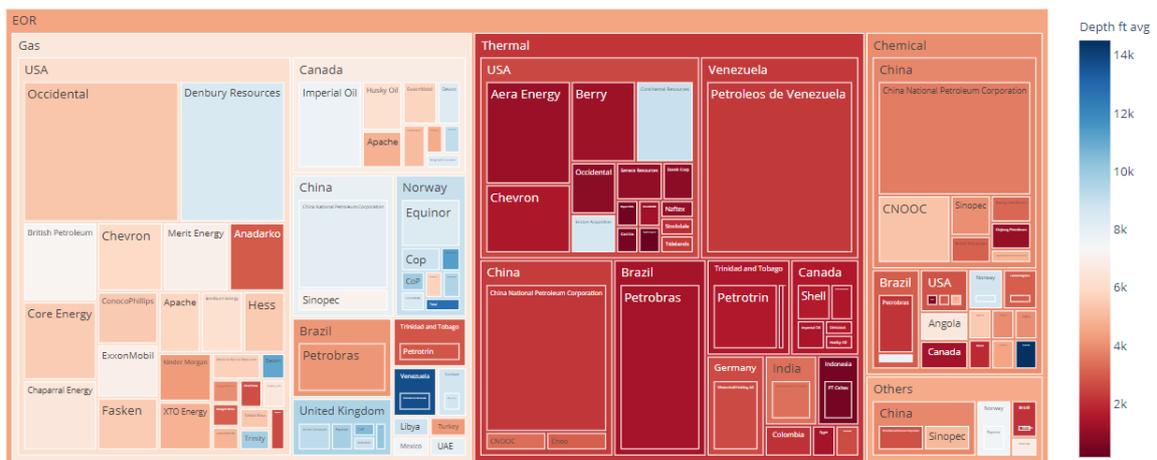


Figura 23. *Treemap* exibindo uma representação hierarquizada dos dados relativos às empresas operadoras dos projetos de EOR, agrupados pelas classes (métodos de injeção de gás, térmicos, químicos e outros) e respectivos países em que estão localizados tais campos. Cada retângulo, representando uma dada empresa, tem seu tamanho proporcional à quantidade de projetos que tal empresa possui. Além disso, as cores dos retângulos, conforme a escala, remetem à profundidade média dos reservatórios em que se teve aplicação de EOR para uma dada classificação.

Os *boxplots* da Figura 24 trazem a distribuição dos parâmetros área e número de poços produtores e injetores de cada campo distribuídos de acordo com as classes de EOR e abrangência da empresa. Observa-se que as pequenas e médias empresas operadoras (classificada como S e M) possuem, em média,

menores áreas e número de poços produtores e injetores em seus campos, especialmente para métodos térmicos e injeção de gás nos quais elas possuem maior atuação; apesar disso, apresentam densidades de poços maiores ou equivalentes às NOCs e IOCs. Ou seja, campos operados por empresas médias e pequenas tendem a empregar maior densidade e menor espaçamento entre poços por meio de estratégias como infill drilling, pertinente principalmente em campos maduros já bastante depletados, com a estratégia de recuperar o óleo remanescente e zonas não varridas. Já as NOCs e IOCs, por sua vez, tendem a ter foco em ativos de grande porte (ativos globais) no início de seu desenvolvimento visando maiores taxas internas de retorno, em detrimento de ativos maduros e ganhos já marginais.

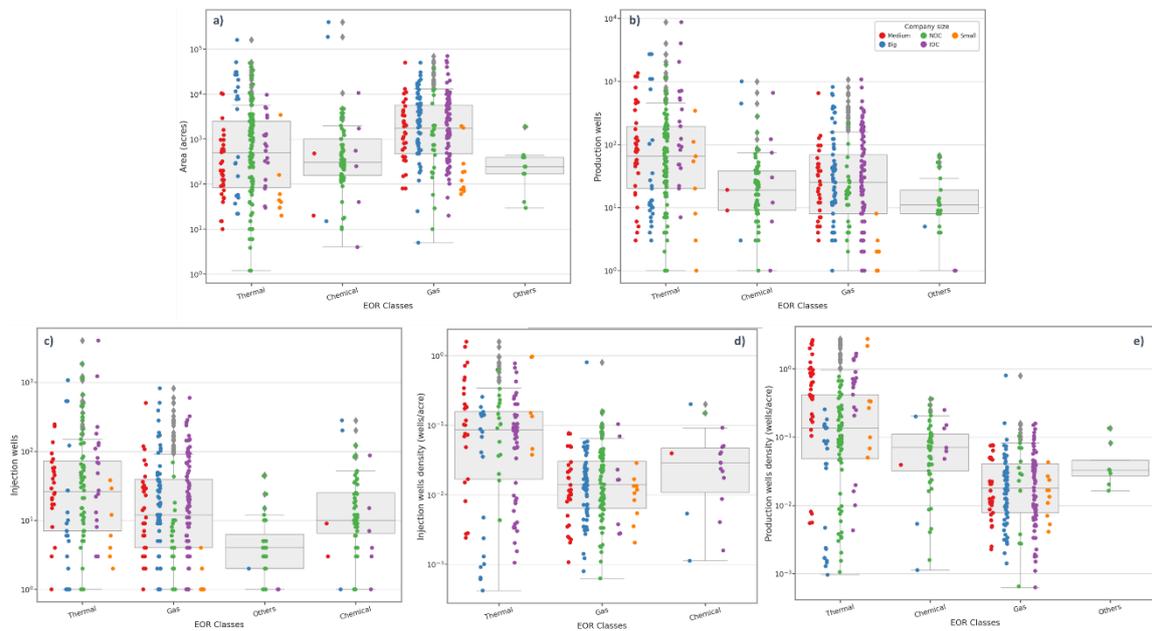


Figura 24. *Boxplots* para as variáveis a) área em acres do projeto, b) e c) número de poços produtores e injetores utilizados, e d) e e) para as respectivas densidades de poços nos projetos de EOR. Os dados são segmentados, em um primeiro nível, em relação às classes de métodos de EOR utilizados. Em um segundo nível, são categorizados quanto ao porte e atuação das companhias operadoras, podendo ser pequenas empresas (Small), médias empresas (Medium), grandes empresas (Big), companhias nacionais e/ou estatais – National Oil Companies (NOCs), e companhias privadas com atuação internacional – International Oil Companies (IOCs).

5.2. Análises por SOM dos métodos de EOR

O treinamento do SOM obteve erros de quantização final e topográfico de 0,193 e 0,075, respectivamente. A Matriz-U resultante da aplicação do algoritmo no conjunto de dados pode ser observada na Figura 25. Observa-se boa dispersão dos conjuntos amostrais de dados pela Matriz-U, o que sugere boa representatividade dos grupos dos dados em questão. Já a Figura 26 ilustra a Matriz-U com os nomes de exemplos de projetos de EOR representados por cada um dos BMUs.

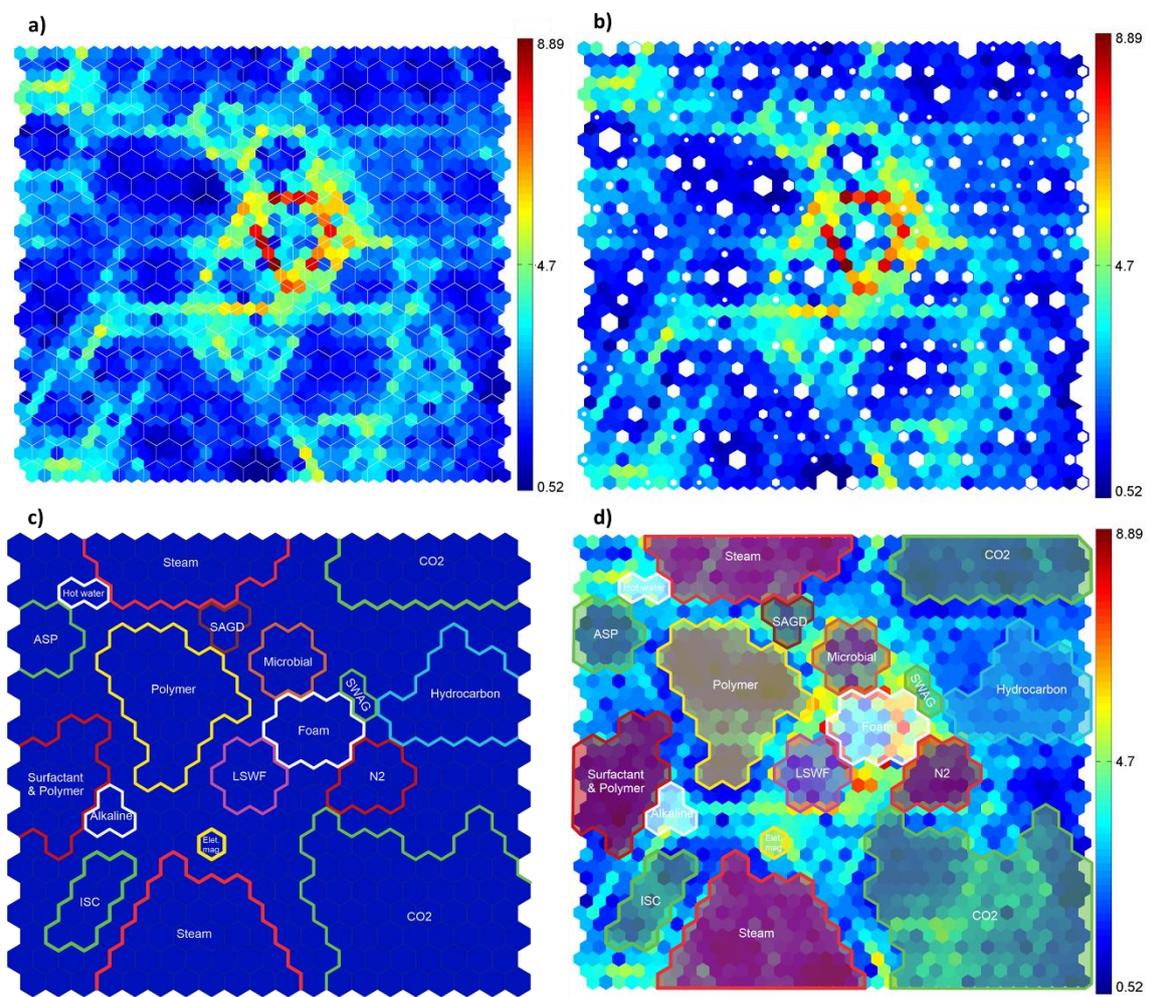


Figura 25. Matriz-U resultante do treinamento realizado com os dados de projetos de EOR, com quatro níveis de destaques: a) Matriz-U com representação da rede hexagonal de neurônios, onde cada contorno branco externo de hexágono representa um neurônio da rede neural artificial, e em que os hexágonos adjacentes à borda indicam a medida de dissimilaridade. As cores mais quentes indicam altas dissimilaridades do neurônio em questão com os seus adjacentes, e as cores mais frias maiores similaridades; b) Matriz-U com hits representados na cor branca. O tamanho do hexágono é proporcional à quantidade de reservatórios que o respectivo BMU está representando; c) linhas interpretadas relacionadas às regiões de neurônios de maior influência exercida pelos dados do respectivo método de EOR; e d) sobreposição das regiões das linhas interpretadas em c) na imagem da Matriz-U de a).

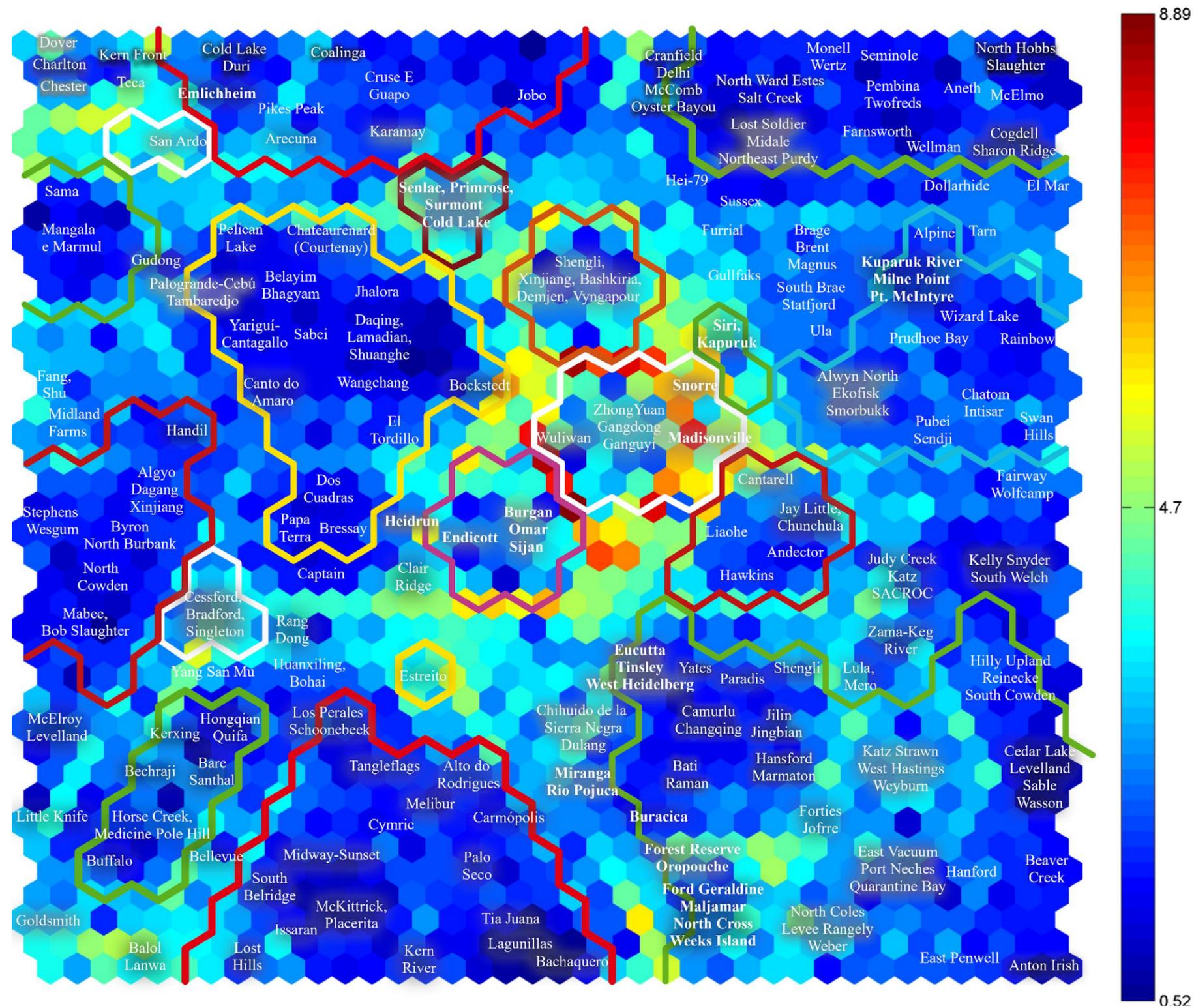


Figura 26. Matriz-U resultante do treinamento realizado com os dados de projetos de EOR. Exemplos de projetos representativos dos respectivos BMUs possuem seus nomes apresentados.

5.2.1. Métodos térmicos

5.2.1.1. Injeção de vapor contínua, cíclica e SAGD

Projetos de injeção de vapor, no inglês *Steam Flooding* (SF) representam 18% da base de dados, com ao todo 230 projetos catalogados, em 16 países. Ainda é o principal método térmico e a maior parte dos projetos encontram-se principalmente nos EUA, Venezuela, China, Brasil e Canadá. Os primeiros projetos catalogados de SF tiveram seu início no final da década de 50 nos EUA, seguido por Venezuela e Canadá. Nas décadas seguintes teve aplicação em maior escala também no Brasil, Trindade e Tobago, China e Índia em reservatórios com óleo viscoso. Na China atualmente cerca de 80% da produção de campos com óleo pesado provém da utilização de injeção de vapor. Verifica-se nos projetos catalogados diversidade em relação a diferentes configurações dos processos de injeção de água e vapor, com a maior parte correspondendo a (i) injeção de vapor contínua, seguida por (ii) injeção cíclica/em batelada de vapor, (iii) injeção de água quente, no inglês Hot Water (HW), e (iv) injeção de vapor em poços direcionais, no inglês Steam Assisted Gravity Drainage (SAGD). Em especial, o Canadá foi o primeiro país a conduzir projeto de SAGD, nos campos de Christina Lake and MacKay River.

Com base nas *component plots* das Figuras 27 e 28, foram analisados os conjuntos de neurônios com maior presença de projetos associados a injeção de vapor. Esses neurônios possuem valores médios de porosidade de 27% (mas com *clusters* apresentando 16 a 37% de porosidade, permeabilidades de em média 1000 mD, profundidades baixas de 3000 ft, °API em sua maioria de até 17, viscosidade acima de 200 cP, e valores estimados médios de garganta de poro acima de 3,3. Via inspeção dos reservatórios presentes em tais neurônios, nota-se *outliers* com permeabilidades da ordem de 10 mD – como a formação Diatomite do campo de South Belridge (MURER et al., 1997). Dentre os projetos de injeção de vapor, os que utilizaram SAGD tiveram aplicação em reservatórios com porosidades médias de 22 a 37%, permeabilidades de 900 a 6000 mD, viscosidades de 1000 a 14000 cP, °API de 9 a 19 e profundidades de 2000 a

3200 ft. Dessa forma, projetos com o uso de SAGD tendem a possuir aplicação em óleos mais pesados e formações menos espessas, devido aos poços direcionais, e mais favoráveis ao fluxo do óleo se comparados aos restantes dos projetos de injeção de vapor.

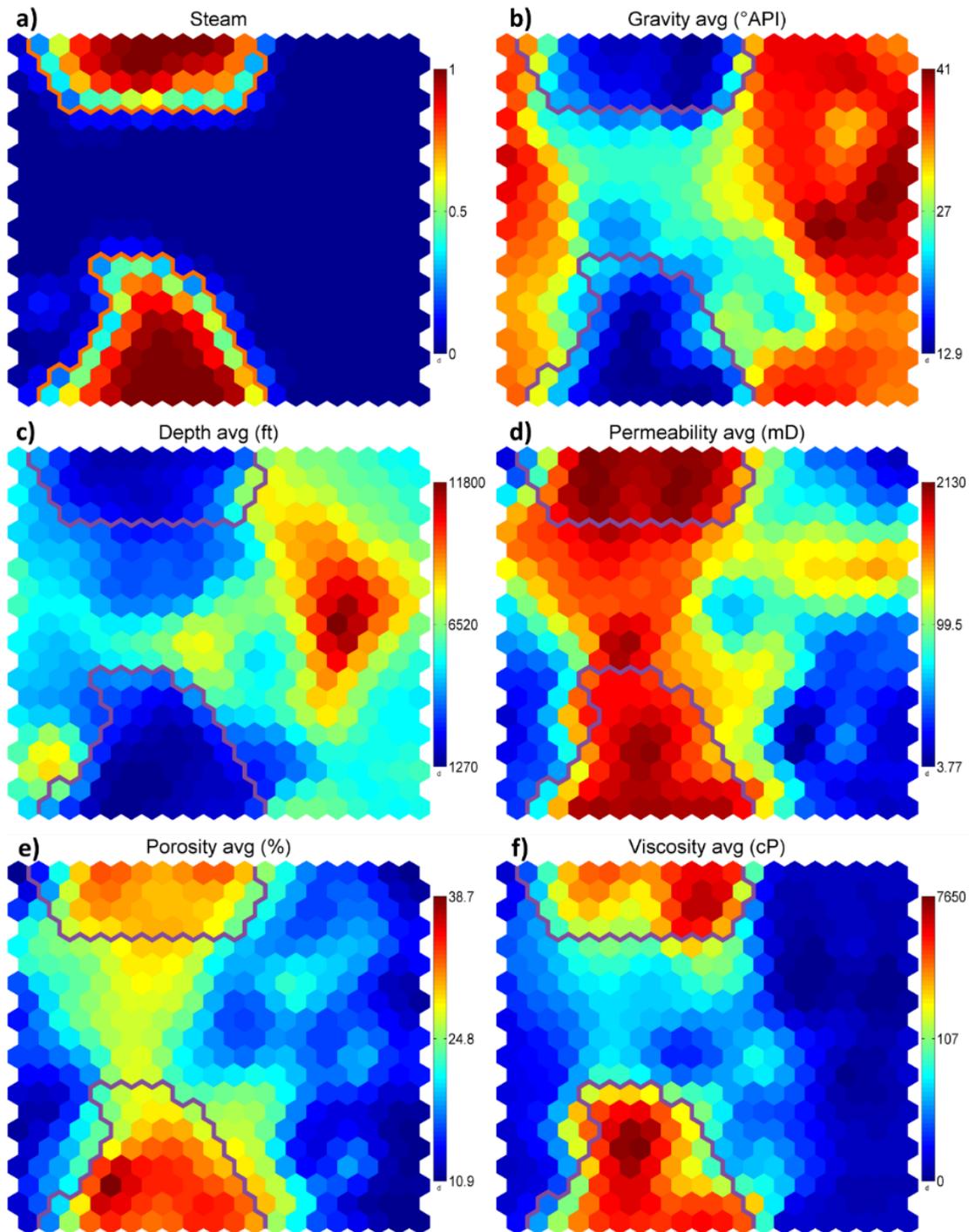


Figura 27. a) *Component plots* referente ao método de injeção de vapor; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor.

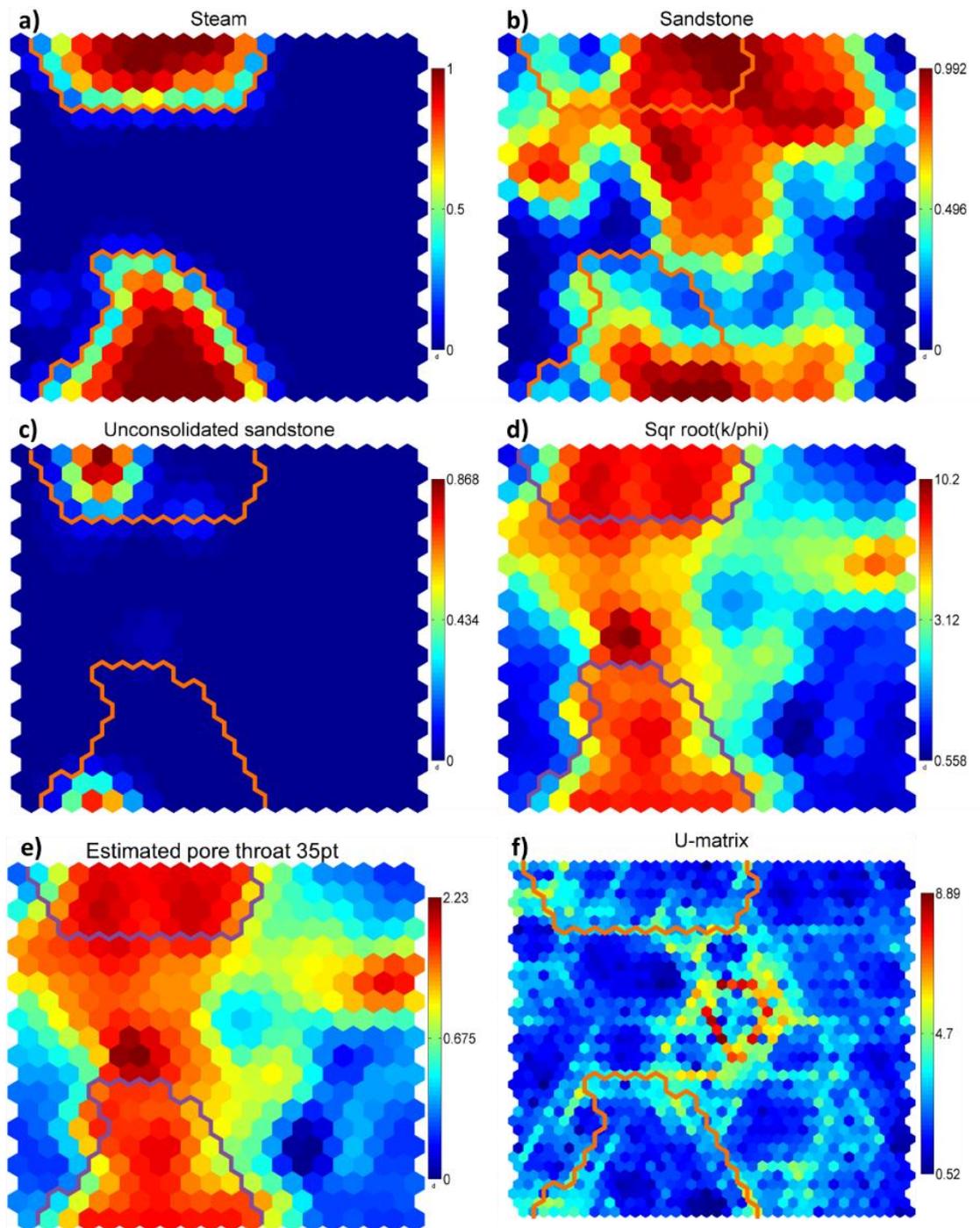


Figura 28. a) *Component plot* referente ao método de injeção de vapor; b) a f) *component plots* de variáveis relacionadas às litologias dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor.

Conforme as *component plots* da Figura 29, os projetos de injeção de vapor tiveram sua aplicação majoritariamente em reservatórios areníticos. Nas 9 aplicações catalogadas em carbonatos, os reservatórios possuem valores menores de permeabilidades (na faixa de 10 a 100 mD) e porosidades (na faixa

de até 20%), mas que são considerados valores altos para formações de matrizes carbonáticas. Além disso, tais projetos de injeção de vapor em carbonatos possuem °API acima de 16, o que é maior se comparados aos projetos de injeção de vapor em arenitos. Dentre os reservatórios areníticos, 38 projetos tiveram aplicação em arenitos inconsolidados, que apresentam médias de valores maiores para porosidade e permeabilidade e profundidades menores se comparados aos reservatórios com arenitos consolidados. Tais tendências de valores menores de °API e valores maiores de propriedades hidráulicas para projetos em reservatórios areníticos também podem ser observadas para os métodos de HWF e SAGD.

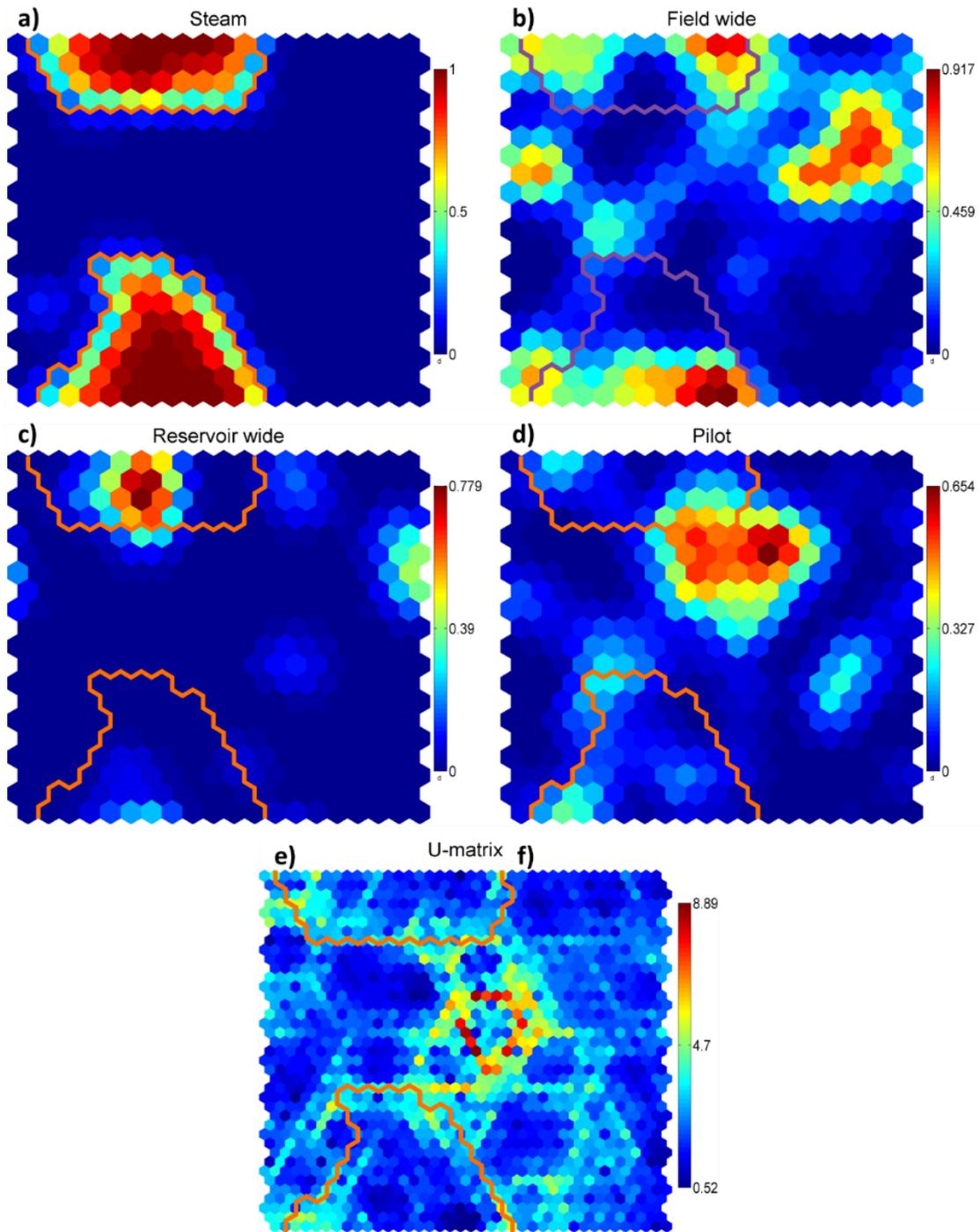


Figura 29. a) *Component plot* referente ao método de injeção de vapor; b) a d) *component plots* referentes à escala do projeto (se piloto, escala de reservatório ou escala de campo); e) a Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor referente a injeção de vapor.

Conforme as component plots da Figura 29, nota-se que grande parte dos projetos de injeção de vapor já possuem aplicação em escala de campo. A injeção de vapor possui aplicação em campos e pesquisa há mais de 50 anos, conforme pode ser visto no gráfico de histórico de projetos e de artigos

publicados das Figuras 3 e 15 e no boxplot da Figura 30. Os primeiros projetos catalogados na base de dados remontam ao final da década de 1950, estando entre os primeiros métodos de EOR a serem aplicados em escala de campo. Para a aplicação de injeção de vapor em geral e combustão in-situ, embora haja flutuações no número de projetos que podem estar associadas à volatilidade do preço do barril conforme Figura 30 (histórico projetos), verifica-se projetos com início de execução em diferentes cenários de preço do barril de petróleo, inclusive períodos com valor ajustado do barril abaixo de 30 dólares. Nos anos 90 observa-se um aumento no número de projetos de injeção de vapor, mesmo com preço do barril baixo, o que pode ser atribuído em parte a programas governamentais de incentivo à adoção de técnicas de recuperação avançada. Já para o SAGD observa-se sua aplicação em cenários com preços do barril mais elevados, o que pode estar associado à complexidade operacional e seus custos mais elevados. Os 5 projetos em maior escala que tiveram aplicação de aquecimento eletromagnético possuem valores baixos de preço do barril no início de sua execução; tais projetos foram executados por grandes empresas, em especial NOCs, que possuem condições robustas de investimento na pesquisa e teste exploratório de novas tecnologias.

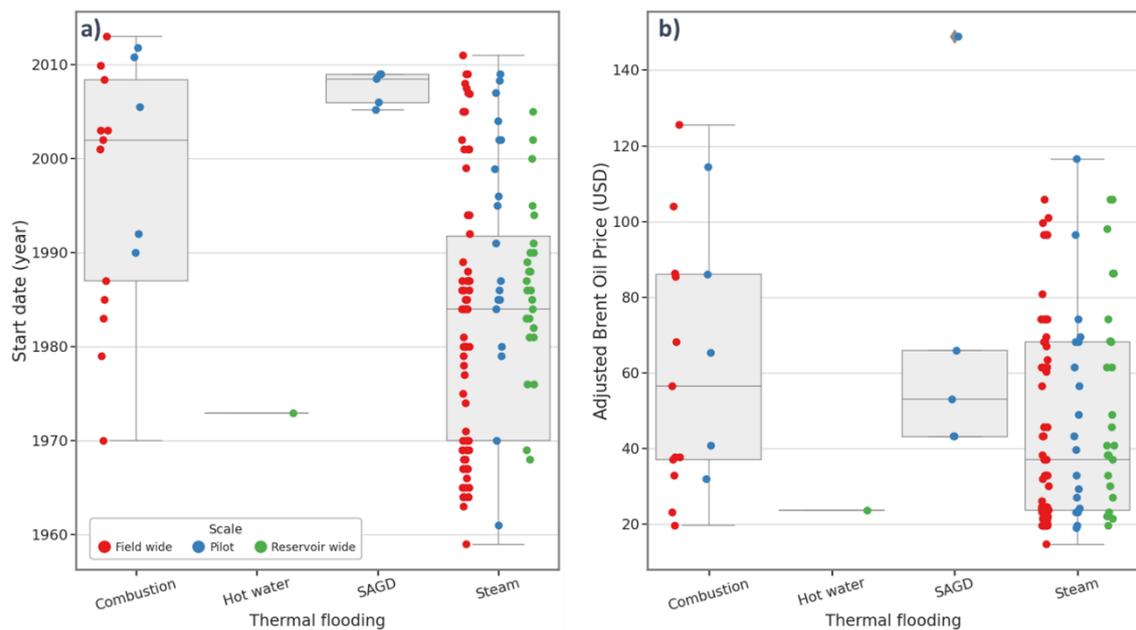


Figura 30. Boxplot referente às variáveis data de início do projeto de EOR e preço do barril do petróleo descontado a inflação para à data de início. Segmentação dos dados quanto ao tipo de método térmico.

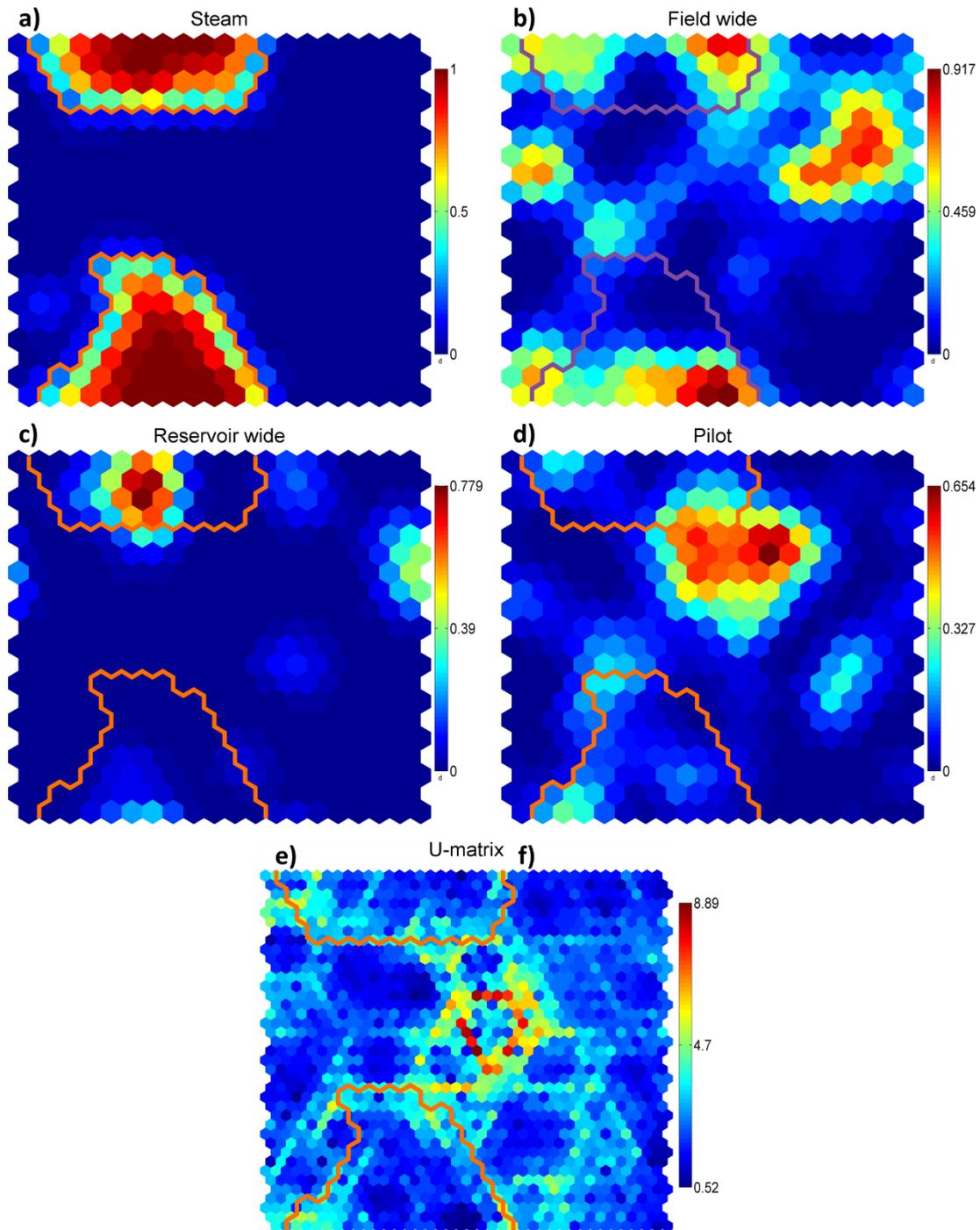


Figura 31. a) *Component plot* referente ao método de injeção de vapor; b) a d) *component plots* referentes à escala do projeto (se piloto, escala de reservatório ou escala de campo); e) a Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de vapor referente a injeção de vapor.

A injeção de vapor é o único método térmico em que foram identificados projetos *offshore*. São 7 projetos localizados, por exemplo, em Bohai Bay, na China, e no Congo. Os projetos *offshore* de injeção de vapor possuem uma média de preço ajustado do barril no início do projeto de 53 USD, enquanto os *onshore* possuem o preço médio do barril a 44 USD.

Conforme os *boxplots* da Figura 31, é possível notar que a maior parte dos projetos de injeção de vapor são implementados em escala de campo, e que esses possuem valores das densidades de poços maiores que projetos pilotos ou em escala de reservatório. Pode-se observar também que empresas pequenas possuem campos com em geral menor área e menores números de poços produtores e injetores. Já NOCs e IOCs possuem grande variabilidade, com campos apresentando tanto baixos quanto altos valores de área e números de poços produtores e injetores. Em relação às densidades de poços verifica-se equivalência entre os diferentes portes de empresas, exceto para NOC que apresenta uma variabilidade maior.

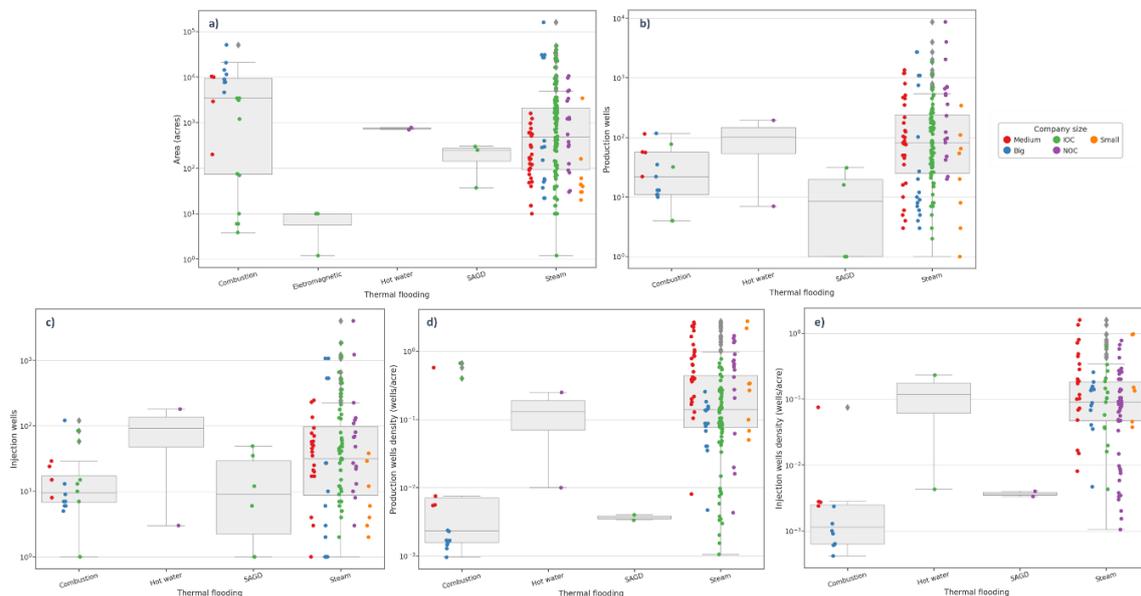


Figura 32. Boxplot referente às variáveis a) área em acres, b) número de poços produtores, c) número de poços injetores, e densidade de poços d) produtores e e) injetores. Segmentação dos dados, num primeiro nível, quanto ao método térmico utilizado para cada boxplot, e quanto ao porte de atuação das empresas operadoras como diferentes cores para os pontos.

Conforme boxplot da Figura 32, em especial NOCs possuem campos de injeção de vapor em profundidades significativamente maiores que o restante das companhias operadoras. Em relação ao °API, dos projetos catalogados as pequenas empresas se caracterizam por aplicar injeção de vapor em reservatórios com °API entre 12 e 20 e as IOC em °API menor que 16; as NOC apresentam maior variabilidade em sua aplicação.

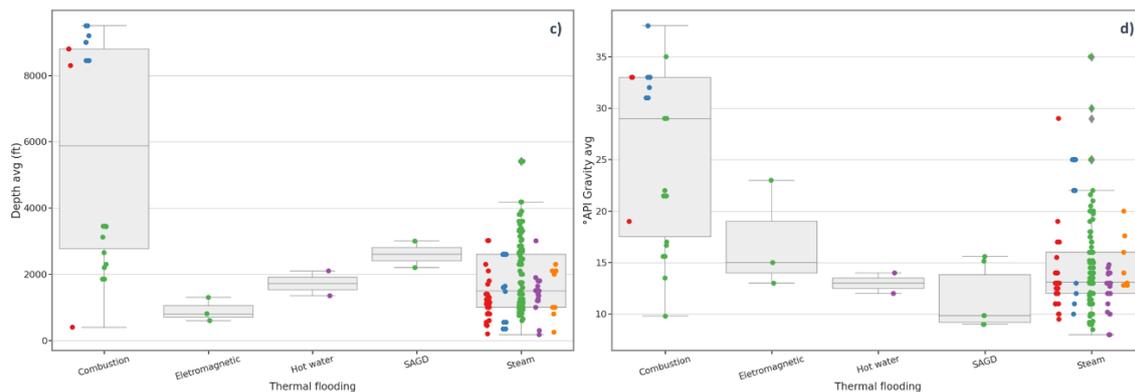


Figura 33. Boxplot referente às variáveis a) profundidade e b) °API. Segmentação dos dados, num primeiro nível, quanto ao método térmico utilizado para cada boxplot, e quanto ao porte de atuação das empresas operadoras como diferentes cores para os pontos.

5.2.1.2. Combustão in-situ

Projetos de combustão in-situ (ISC) representam 6% da base de dados, com ao todo 91 projetos catalogados, em 7 países. A maior parte dos projetos encontra-se principalmente nos EUA e China. Os primeiros projetos catalogados tiveram seu início no final da década de 60 nos EUA e Brasil, com ampla aplicação em maior escala nas décadas seguintes também para a Índia, China e Venezuela e outros países. A maior parte dos projetos trataram-se de aplicações em escala de campo conduzidas por empresas de grande porte, como NOCs e IOCs.

Com base nas *component plots* das Figuras 34 e 35, pode-se verificar dois agrupamentos principais de neurônios com reservatórios associados a projetos com ISC: (i) com reservatórios areníticos e (ii) relacionado com reservatórios carbonáticos. O grupo de neurônios (i) apresenta valores de porosidade próximos a 25, permeabilidades da ordem de 1000 mD, profundidade de cerca de 3500 ft, °API próximo a 16 e viscosidade de 100 cP; porém, apesar de o grupo (i) ser constituído majoritariamente de aplicações em arenitos, quando comparado com processos de injeção de vapor verifica-se pouca ocorrência de utilização de ISC em arenitos não consolidados. O grupo (ii) se mostra um caso particular, estando associado principalmente a reservatórios carbonáticos com °API mais leve, profundidades mais altas e baixas permeabilidades, característico de formações geológicas como Red River A, B e C (KUMAR; FASSIHI, 1995; MILLER, 1995). Além disso, o grupo (ii) apresenta densidade de poços produtores e injetores menor em relação ao grupo (i). Se comparada

aos demais métodos térmicos, em especial injeção de vapor, nota-se que em geral a aplicação da ISC ocorre em formações geológicas com maiores intervalos de profundidade, satisfazendo condições mínimas para injeção de ar, além de menores permeabilidades, porosidades, gargantas estimadas de poro e espessuras de reservatório.

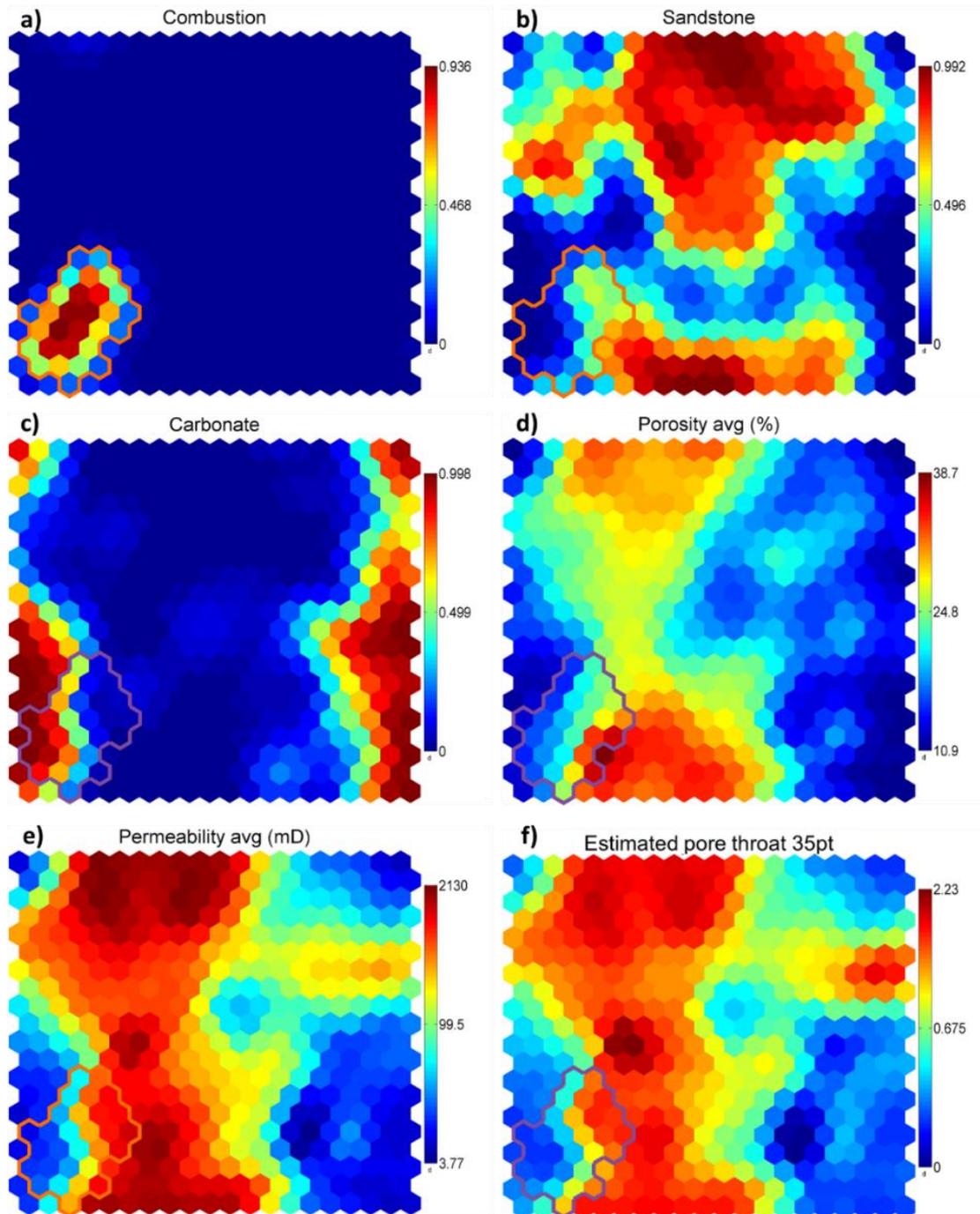


Figura 34. a) Component plot referente ao método combustão in-situ; b) a f) component plots referentes à escala do projeto (se piloto, escala de reservatório ou escala de campo). A linha em destaque delimita os neurônios ativados por reservatórios associados ao método de combustão in-situ.

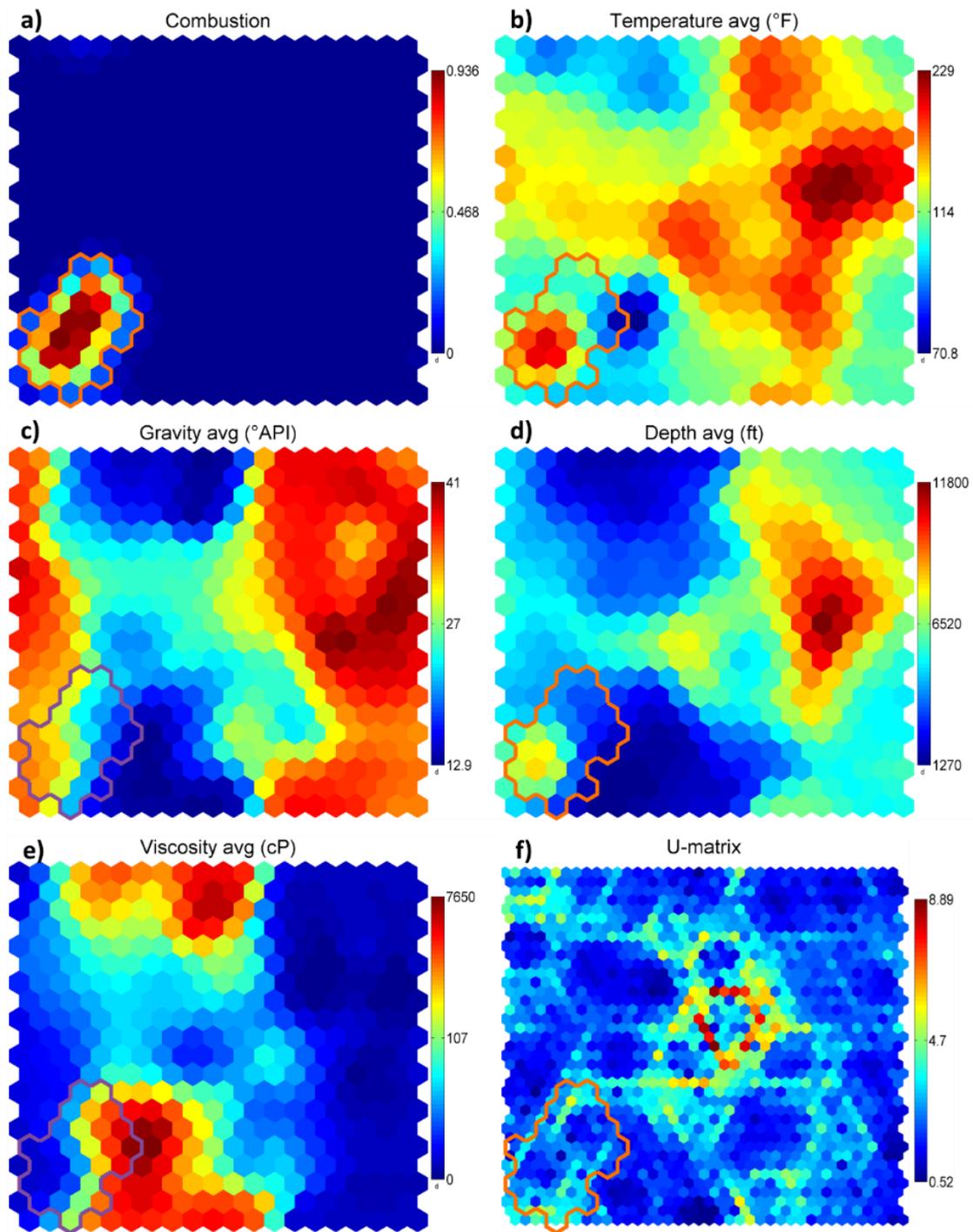


Figura 35. a) *Component plot* referente ao método de combustão *in-situ*; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de combustão *in-situ*.

5.2.2. Injeção de gases

5.2.2.1. Hidrocarbonetos

Projetos de injeção de gás hidrocarboneto (HC) representam 9% da base de dados, com ao todo 142 projetos catalogados, em 16 países. A maior parte dos projetos encontram-se principalmente nos EUA e Canadá, com seu início no final dos anos 1960, com maior aplicação em larga escala nos anos seguintes na região do Mar do Norte e China.

Com base nas *component plots* da Figuras 36, pode-se verificar que os valores médios para as variáveis dos conjuntos de *clusters* de maior representatividade dos projetos de injeção de CO₂ possuem valores médios de porosidade de 14%, permeabilidades de 90 mD, profundidade de 9000 ft, temperatura de 200°F, °API de 36, viscosidade de 0.8 cP, valores estimados médios de garganta de poro de 3,4 e de mobilidade do óleo na ordem de 200, e preço ajustado do barril no início do projeto de 45 USD.

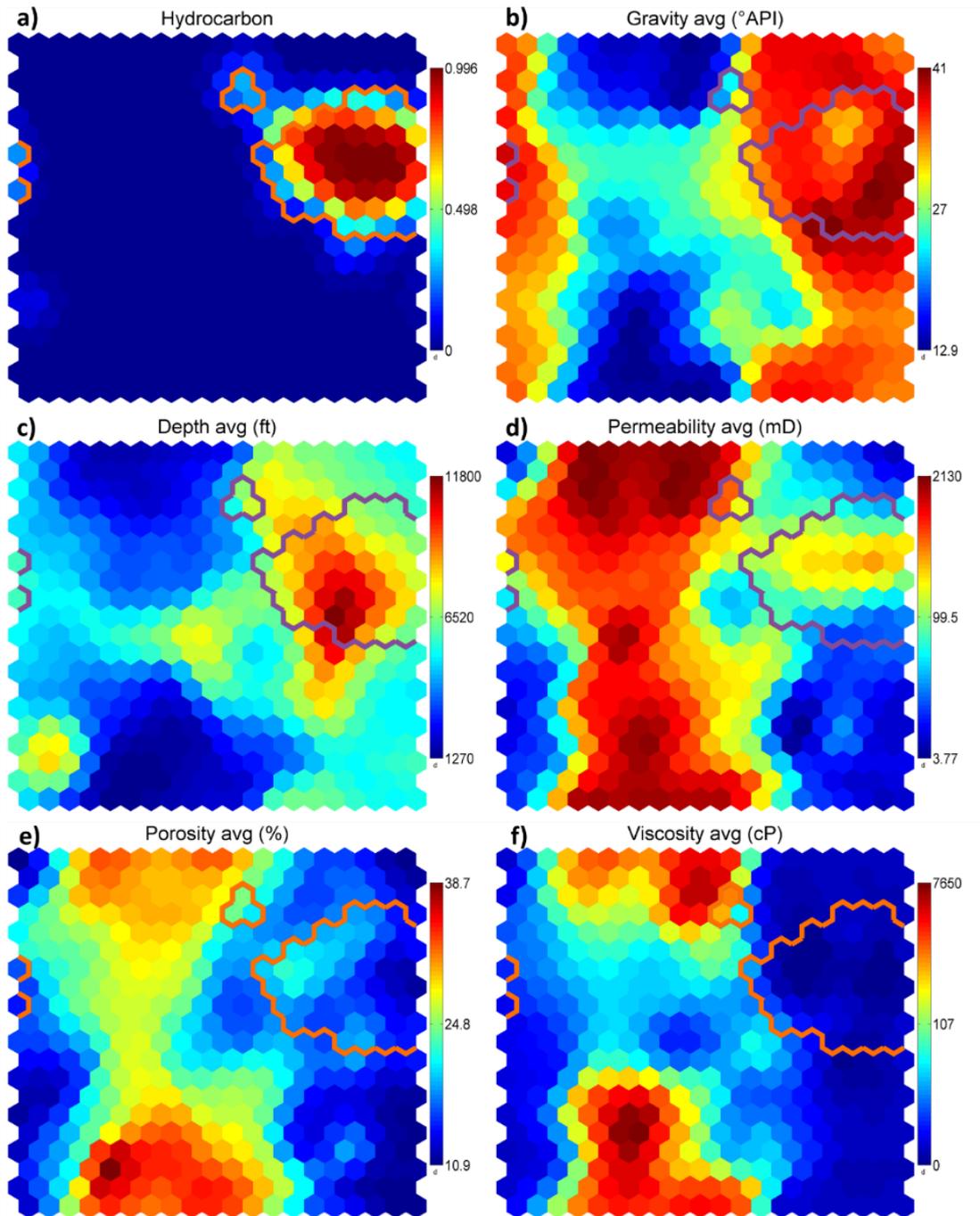


Figura 36. a) *Component plot* referente ao método de injeção de gás hidrocarboneto; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás hidrocarboneto.

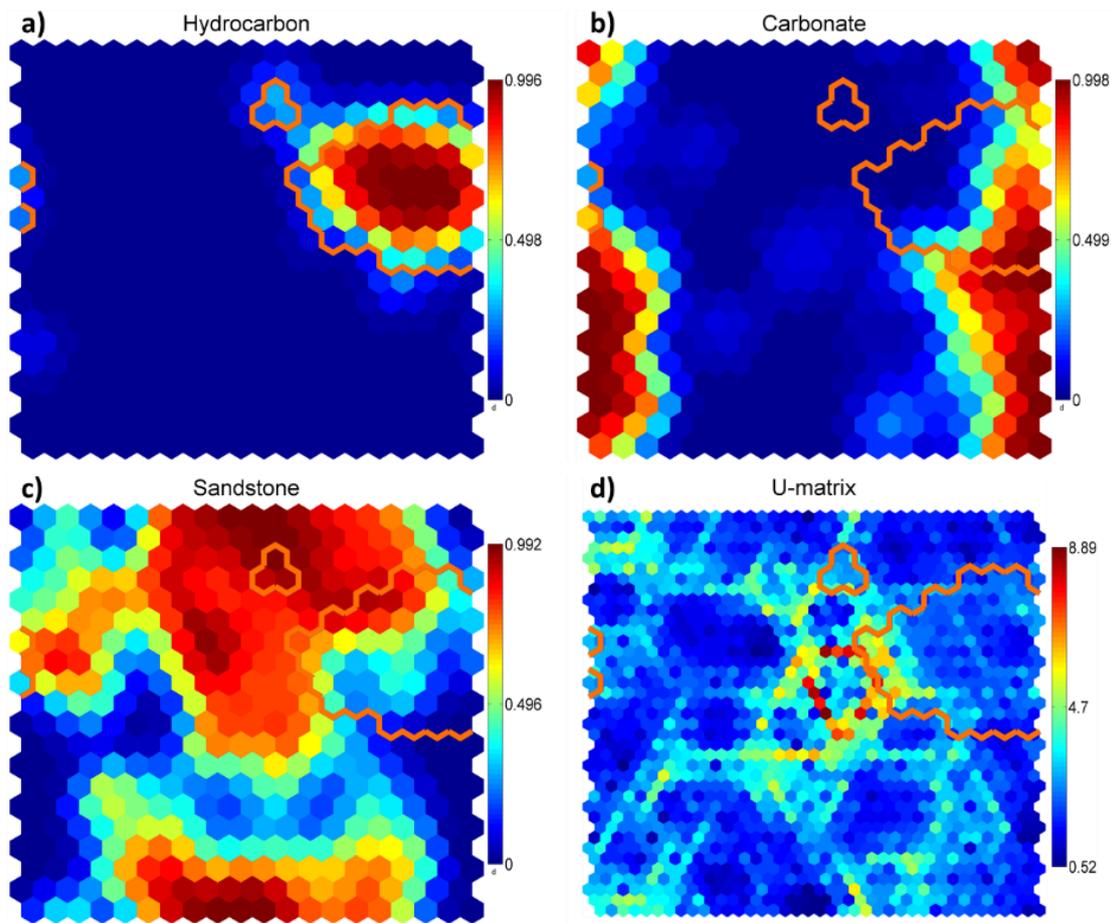


Figura 37. a) *Component plot* referente ao método de injeção de gás hidrocarboneto; b) e c) *component plots* de variáveis relacionadas às litologias dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás hidrocarboneto.

Os projetos de injeção de HC foram aplicados tanto em reservatórios areníticos como em dolomíticos e calcário, conforme as *component plots* da Figura 37. Para os projetos de injeção de HC, quando se comparado os reservatórios areníticos aos carbonáticos, os reservatórios areníticos possuem maior média de porosidade e permeabilidade, menor ρ_{API} , e 22 projetos com profundidade acima de 10000 ft. Já os carbonáticos que tiveram injeção de HC possuem profundidades de até 10000 ft, com exceção de dois projetos com profundidades acima dos 15000 ft. A injeção de HC é o método de EOR com maior número *offshore* de projetos, havendo 52 projetos *offshore* catalogados, dos quais 15 são projetos com escala de campo e localizados no Mar do Norte.

Os primeiros projetos catalogados na base de dados remontam ao final da década de 1960, havendo uma relação, ainda que fraca, com o preço do barril do petróleo. Conforme o boxplot da Figura 38 abaixo, nos anos 90 observa-se

um aumento no número de projetos, mesmo com preço do barril baixo, o que pode ser atribuído em parte a na época o programa do DOE USA de incentivo à adoção de técnicas de recuperação avançada.

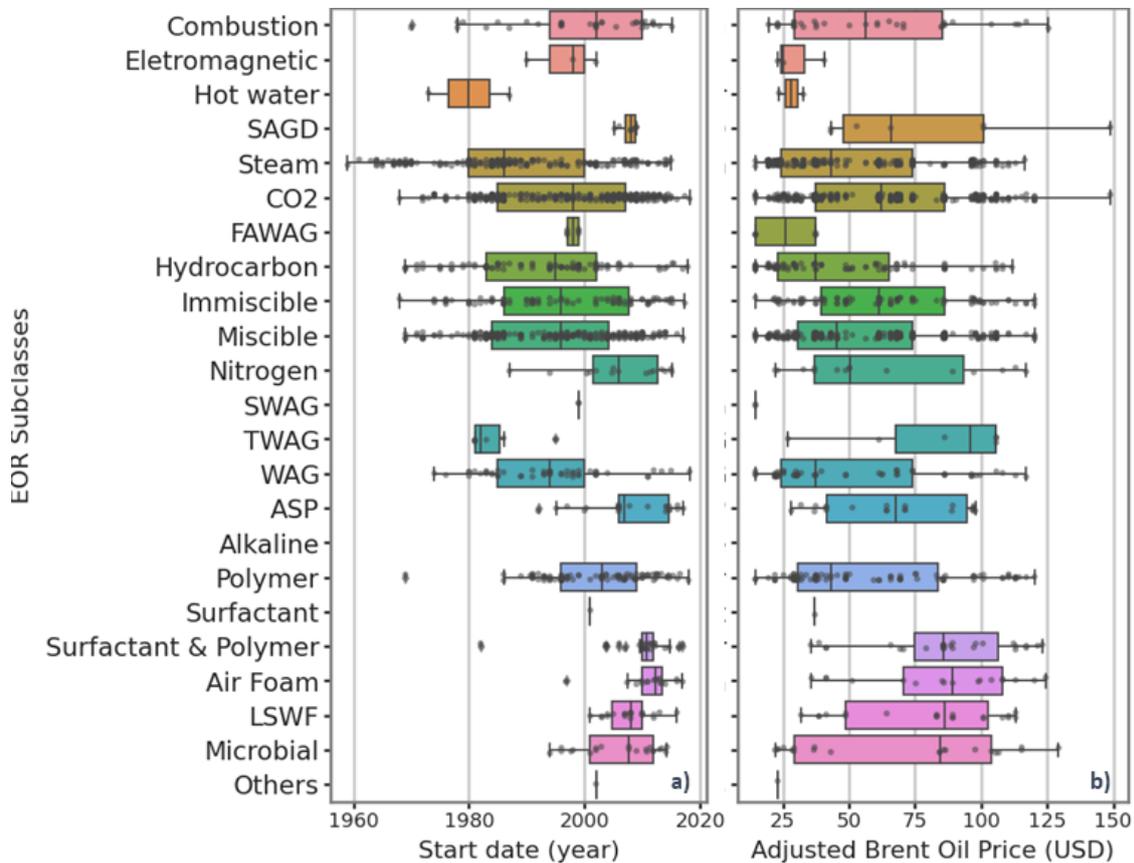


Figura 38. *Boxplots* referente às variáveis data de início do projeto de EOR e preço do barril do petróleo descontado a inflação para à data de início. Segmentação dos dados quanto às subclasses de métodos de EOR utilizadas.

Se comparada à injeção de CO₂, nota-se a diferença de que 55% dos projetos de injeção de HC tiveram seu início em cenários com preço ajustado do barril inferior a 40 dólares, enquanto para a injeção de CO₂ 68% dos projetos tiveram seu início com barril acima de 40 USD. Além disso, a injeção de HC possui a segunda menor média e terceira menor mediana de valores do preço do barril ajustado no momento do início dos projetos. Tal fato pode ser em parte associado ao motivo de a preços mais altos do barril haver uma priorização para se distribuir e comercializar o gás, monetizando-o, em detrimento de injetá-lo. Além disso, com base nas *component plots* da Figura 39, observa-se que a maior parte dos neurônios e projetos associados a injeção de HC, que em sua maioria

estão associados a valores menores do preço ajustado do barril, possuem valores maiores da variável referente a escala de campo e menor influência de projetos pilotos.

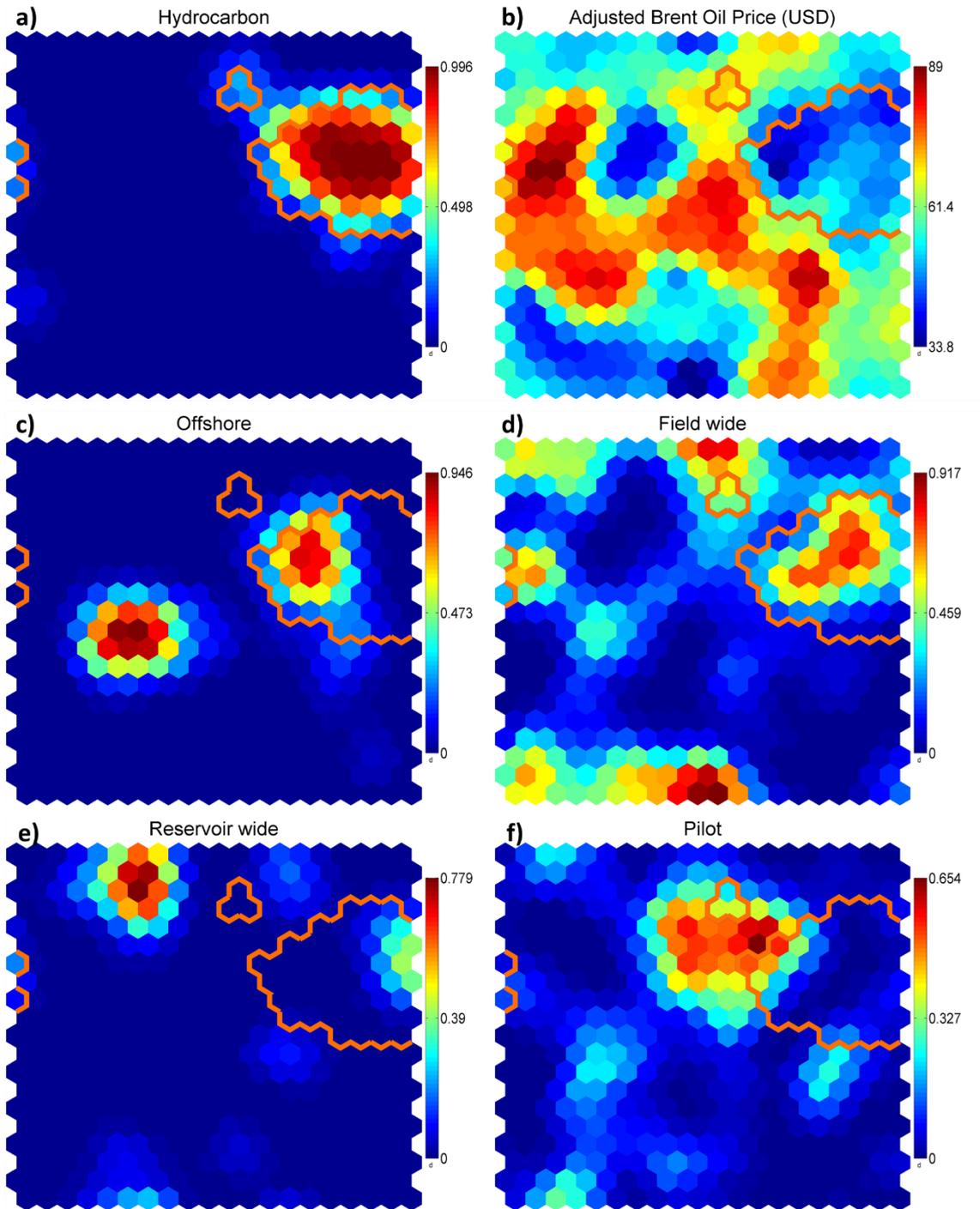


Figura 39. a) *Component plot* referente ao método de gás hidrocarboneto; b) o preço do barril ajustado pela inflação no momento de início dos projetos; c) se o projeto é *offshore*; e) a f) *component plots* referentes à escala do projeto (se implementado na forma de projeto piloto, em escala de reservatório ou em escala de campo).

A injeção de HC é o método de injeção de gás com maior número *offshore* de projetos, havendo 52 projetos *offshore* catalogados, dos quais 15 são projetos com escala de campo e localizados no Mar do Norte. Os projetos *offshore* de injeção de HC possuem uma média de preço ajustado do barril no início do projeto de 53 USD, enquanto os *onshore* possuem o preço médio do barril a 44 USD, conforme é possível observar nas *component plots* da Figura 39.

Pode-se observar nas *component plots* da Figura 40 que para injeção de HC há a predominância de empresas integradas, com poucas empresas de menor porte ou que sejam focadas na E&P. A maior parte dos neurônios de injeção de HC encontram-se fortemente associadas à atuação de empresas IOCs, com menor participação de NOC e Bigs.

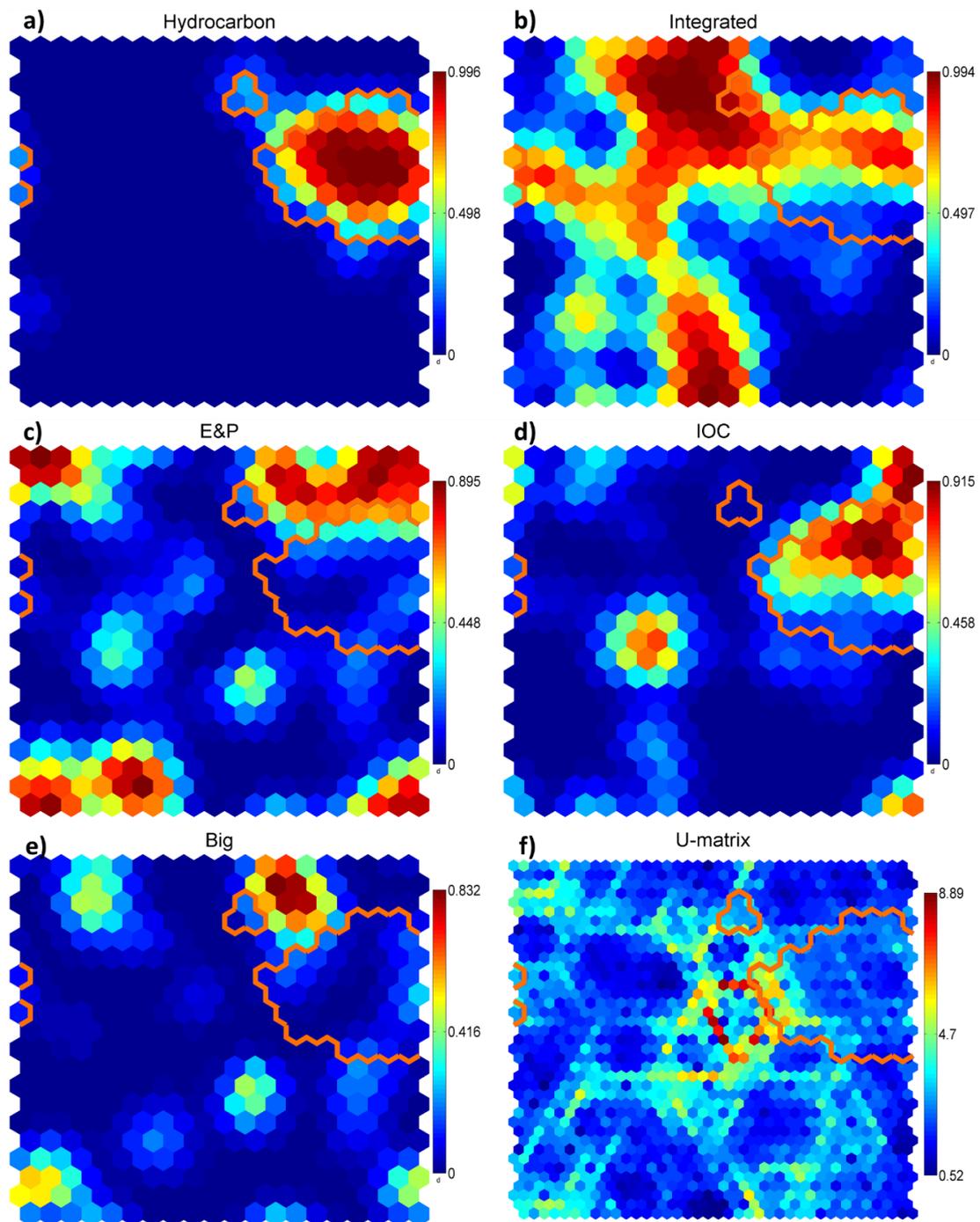


Figura 40. a) *Component plot* referente ao método de injeção de gás hidrocarboneto; b) e c) *component plots* referentes ao segmento de atuação (se integrada na cadeia de valor ou independente e focada na exploração e produção – *Upstream*); d) e e) tamanho da companhia operadora, se International Oil Company (IOC) ou empresa de grande porte, mas com atuação regional; f) Matriz-U.

5.2.2.2. Dióxido de carbono

Projetos de injeção de CO₂ representam 32% da base de dados, com ao todo 510 projetos catalogados, em 15 países. É atualmente o método de EOR com

maior número de projetos. Há uma grande variedade de configurações de processos de injeção de CO₂, tendo variações em relação ao processo ser miscível ou imiscível, e contínuo, alternado com água ou simultâneo. A maior parte dos projetos encontram-se principalmente nos EUA, Canadá, China e Brasil, com menor número de projetos em países como Trinidad Tobago, Turquia, Mar do Norte, Argentina, Hungria, Itália, Malásia e Qatar. Nos EUA a aplicação de CO₂ começou nos anos 1960, com ampla aplicação em larga escala nas décadas seguintes. No Canadá, Trinidad e Tobago, Brasil e Turquia, ainda que em menor número, tiveram seus primeiros projetos de injeção de CO₂ nos anos 1970 e 1980, totalizando 36 projetos nesse período para esses quatro países. Já na China um número significativo de projetos de injeção de CO₂ começou a surgir apenas a partir de 2005. Enquanto até 2010 quase 70% dos projetos de injeção de CO₂ foram aplicados nos EUA, esse número reduziu para 40% até o presente momento, com o crescimento da sua aplicação em outros países.

Com base nas *component plots* da Figura 41, os projetos com processos miscíveis de CO₂ possuem limite inferior de profundidade um pouco maior do que os referentes a processos imiscíveis, necessário para se manter abaixo da pressão de fratura. Projetos imiscíveis também são aplicados em profundidades maiores, porém com pressão de injeção que não é a suficiente para promover miscibilidade. Contudo, a diferença entre os processos miscíveis e imiscíveis de CO₂ se destaca em relação ao °API médio em que são aplicados. Projetos com processos imiscíveis apresentam óleo com amplo range de viscosidade e °API, variando de 11 a 40, enquanto os projetos com processos miscíveis se concentram de 27 a 45 °API.

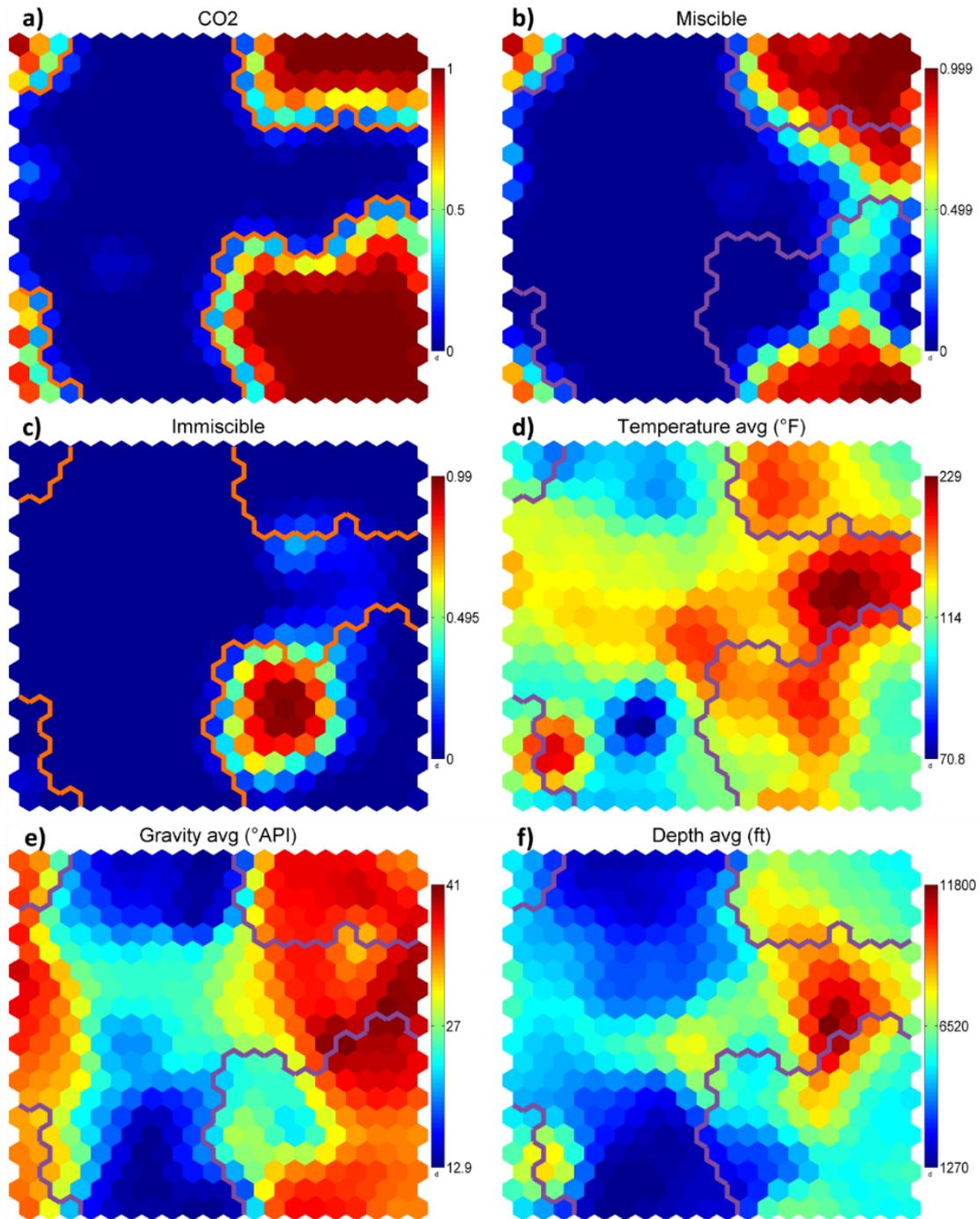


Figura 41. a) *Component plot* referente ao método de injeção de gás CO₂; b) e c) *component plots* relacionadas a se o processo de injeção é miscível ou imiscível; d) a f) *component plots* de variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás CO₂.

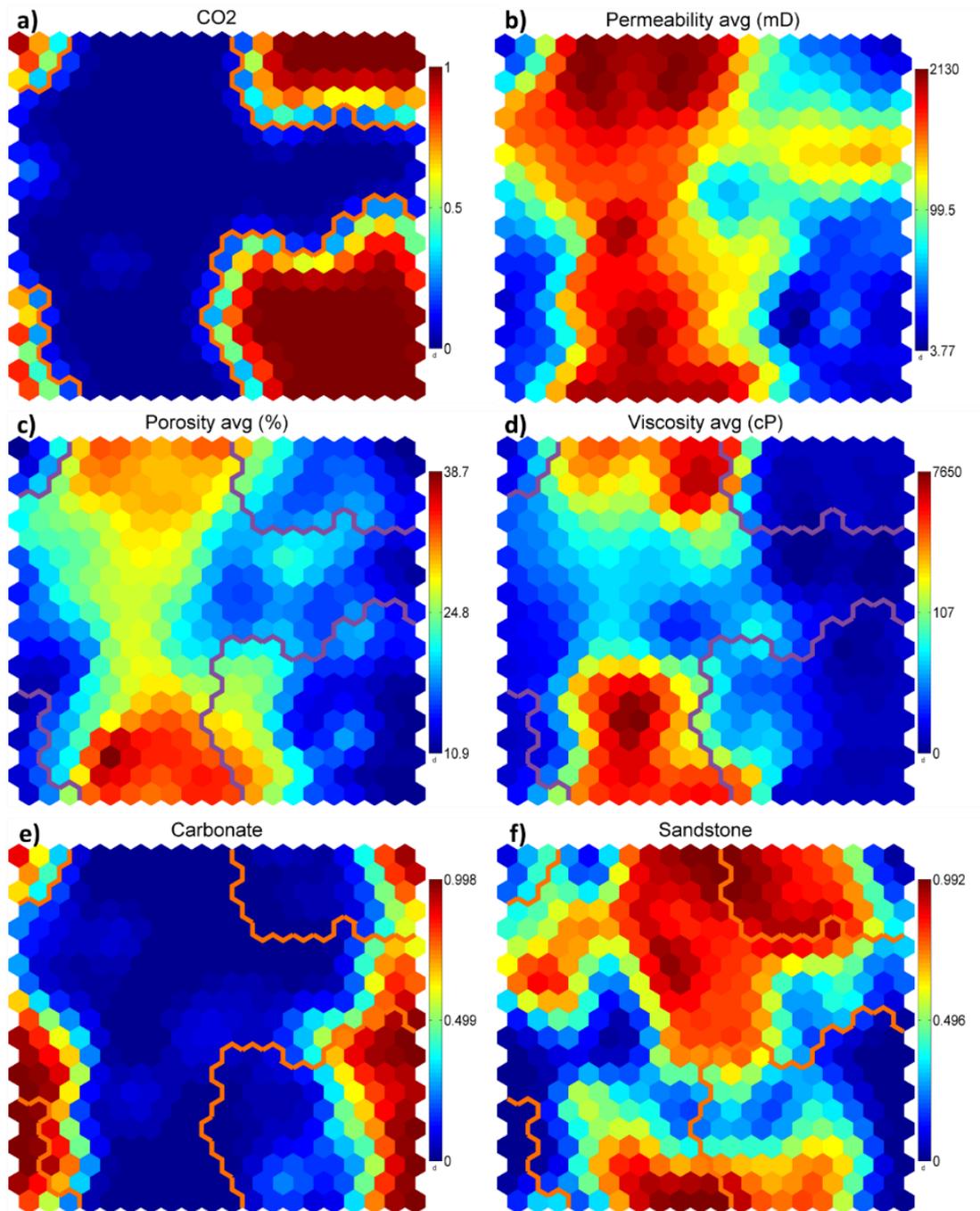


Figura 42. a) *Component plot* referente ao método de injeção de gás CO₂; b) a d) *component plots* de variáveis relacionadas a variáveis básicas dos reservatórios; e) e f) *component plots* de variáveis relacionadas às litologias dominantes dos reservatórios. A linha em destaque delimita os neurônios ativado pelo método de injeção gás CO₂.

Conforme Figura 42, os conjuntos de neurônios com maior quantidade de projetos de injeção de CO₂ possuem valores médios de porosidade de 16%, permeabilidades de 30 mD, profundidade de 5600 ft, temperatura de 145°F, °API de 33, viscosidade de 14 cP, valores estimados altos de garganta de poro (entre 3 e 5,5) e de transmissibilidade, com preço médio ajustado do barril no início do

projeto de 58 USD. Os projetos de injeção de CO₂ foram aplicados tanto em reservatórios areníticos como em dolomíticos e calcários.

A injeção de CO₂ possui aplicação em campos e pesquisa há mais de 50 anos, conforme pode ser visto nos gráficos de histórico de projetos e de artidos nas Figuras 3 e 15. Em especial, com a elevação dos preços do barril entre 1973 e 1985, teve-se um grande aumento no número de projetos de CO₂, especialmente no íterim de 1975 a 1985. O número de projetos de CO₂ voltou a ter um aumento em meados da década de 90, que consistiu em cenário de baixo preço do barril, mas que teve diversos estímulos governamentais, como do Departamento de Energia dos EUA, visando segurança energética e melhor aproveitamento das reservas e a construção de infraestrutura de captura e escoamento de CO₂. Posteriormente, nos anos 2000, com nova elevação do preço do barril e estímulos relacionados a captura e armazenamento de carbono, o número de projetos de injeção de CO₂ volta a aumentar. Conforme pode-se observar no boxplot b) da Figura 38, a injeção de CO₂ possui mediana e média maiores de preço ajustado do barril no início de seus projetos quando se comparada à injeção de gás hidrocarboneto.

Conforme é possível observar nas *component plots* da Figura 43 para a injeção de CO₂ há uma diversidade grande em relação aos segmentos e portes de empresas que lideraram tais projetos, havendo não apenas majors IOC, NOC e integradas no setor de energia, mas também grande participação de empresas independentes focadas no setor de E&P. Mais de dois terços dos projetos de injeção de CO₂ executados por empresas independentes pequenas e focadas na E&P ocorreram após os anos 2000. A injeção de CO₂ se diferencia dentre os métodos de injeção de gás por uma parte significativa dos *clusters* referentes a projetos de EOR estarem associados a empresas de menor porte e de menor receita, especialmente nos EUA.

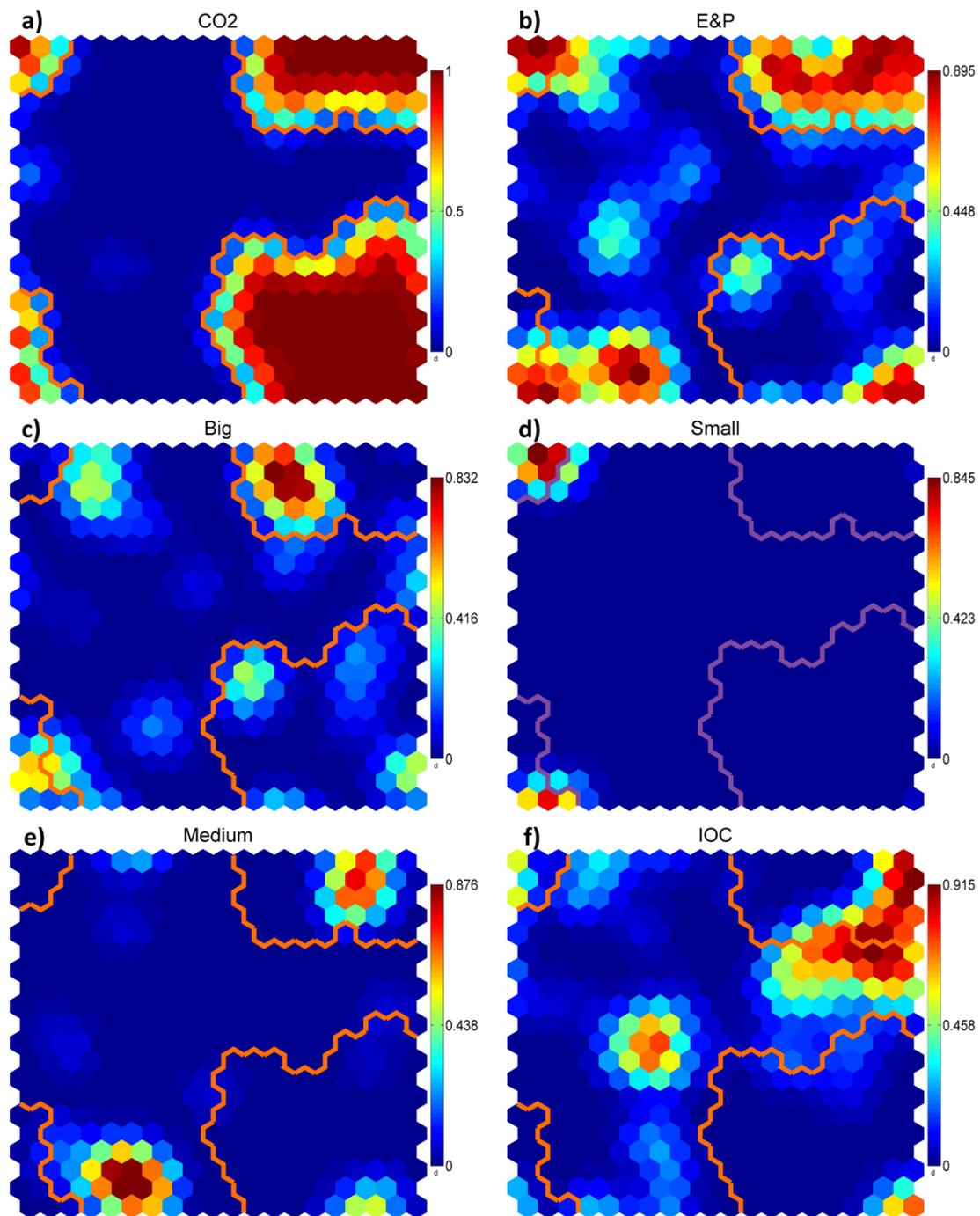


Figura 43. a) *Component plot* referente ao método de injeção de gás CO₂; b) *component plot* referente ao segmento de atuação (se independente e focada na exploração e produção – *Upstream*); c) a f) tamanho da companhia operadora (se empresa de grande porte, mas com atuação regional; se empresas de pequeno e médio porte. e se International Oil Company (IOC)).

5.2.2.3. Nitrogênio

Projetos de injeção de N₂ representam 2,7% da base de dados, com ao todo 37 projetos catalogados, em 9 países. A maior parte dos projetos encontram-se

principalmente nos EUA e China, com projetos pontuais no Canadá, México, Brasil, Trinidad and Tobago e oriente médio. Os primeiros projetos catalogados tiveram seu início no final da década de 1980 nos EUA. Na China o primeiro projeto reportado de injeção de N₂ foi em 2004, tendo um crescimento de sua adoção no país desde então.

Apesar de o número de projetos obtidos de injeção de N₂ ser pequeno, as *component plots* da Figura 44 mostram uma grande variabilidade do range de aplicação. Esse grande range de variabilidade pode ser explicado em parte por conta de 27 dos projetos de injeção de N₂ utilizarem processos imiscíveis, com apenas 5 projetos de injeção miscível. A aplicação de processos imiscíveis possui menor grau de exigência em relação ao cumprimento de requisitos como °API e profundidades mínimas quando se comparado a miscíveis, que visam promover as condições necessárias para miscibilidade do óleo com o gás injetado. Em especial, nota-se que enquanto os projetos imiscíveis de injeção de N₂ possuem aplicação em diversas profundidades, °API e viscosidades de óleo, os miscíveis possuem aplicação em profundidades e °API maiores e viscosidades estritamente menores. Os projetos de injeção de N₂ possuem uma mediana de profundidades maior frente aos projetos de CO₂. Nos projetos de N₂ a viscosidade média do óleo dos reservatórios é a maior dentre os métodos de gás, possivelmente associado ao fato de serem projetos com processo imiscível, estando mais focado no aumento de pressão do reservatório e deslocamento mecânico do óleo do que em efeitos de miscibilidade.

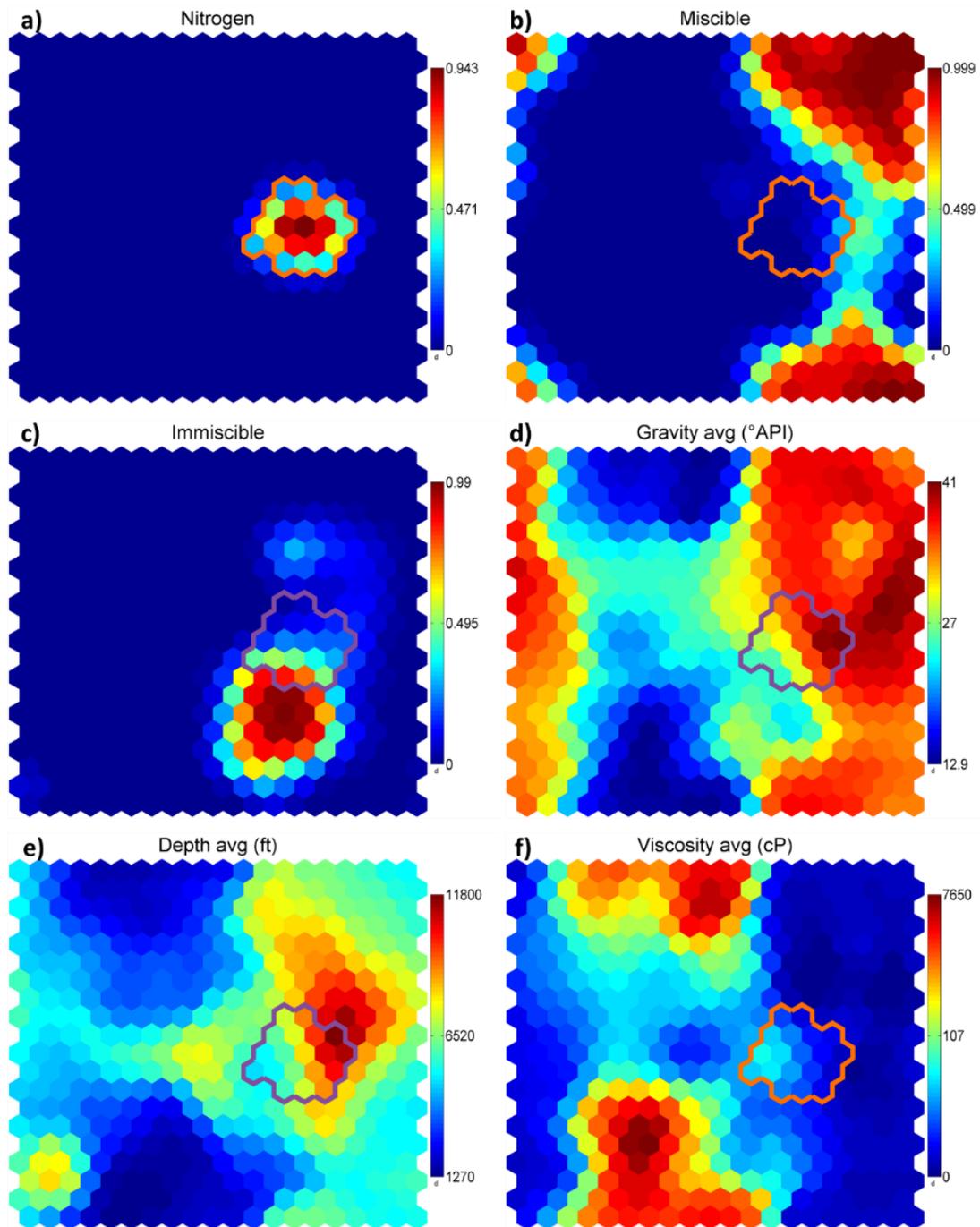


Figura 44. a) *Component plot* referente ao método de injeção de gás nitrogênio; b) e c) em relação a se o processo de injeção do gás é miscível ou imiscível; d) a f) *component plots* de variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás nitrogênio.

Com base ainda nas *component plots* da Figura 44, pode-se verificar que os valores médios para as variáveis dos conjuntos de neurônios de maior representatividade de projetos de injeção de N₂ possuem valores médios de porosidade da ordem de 15%, permeabilidades de 50 mD, profundidade de 8000 ft, temperatura de 200°F, °API de 38, viscosidade de 10 cP, valores estimados

médios de garganta de poro de 3 e de mobilidade do óleo na ordem de 150, preço ajustado do barril no início do projeto de 40 USD. Conforme Figura 45, os projetos de injeção de N2 foram aplicados predominantemente em reservatórios carbonáticos, com apenas cinco casos em reservatórios areníticos. Os projetos de injeção de N2 em carbonatos tiveram °API variando de 30 a 53, profundidade de 7500 a 18000 ft. Já os projetos de injeção de N2 em reservatórios areníticos se concentraram nas faixas de °API 16 a 25 e profundidade na 3000 a 4600 ft.

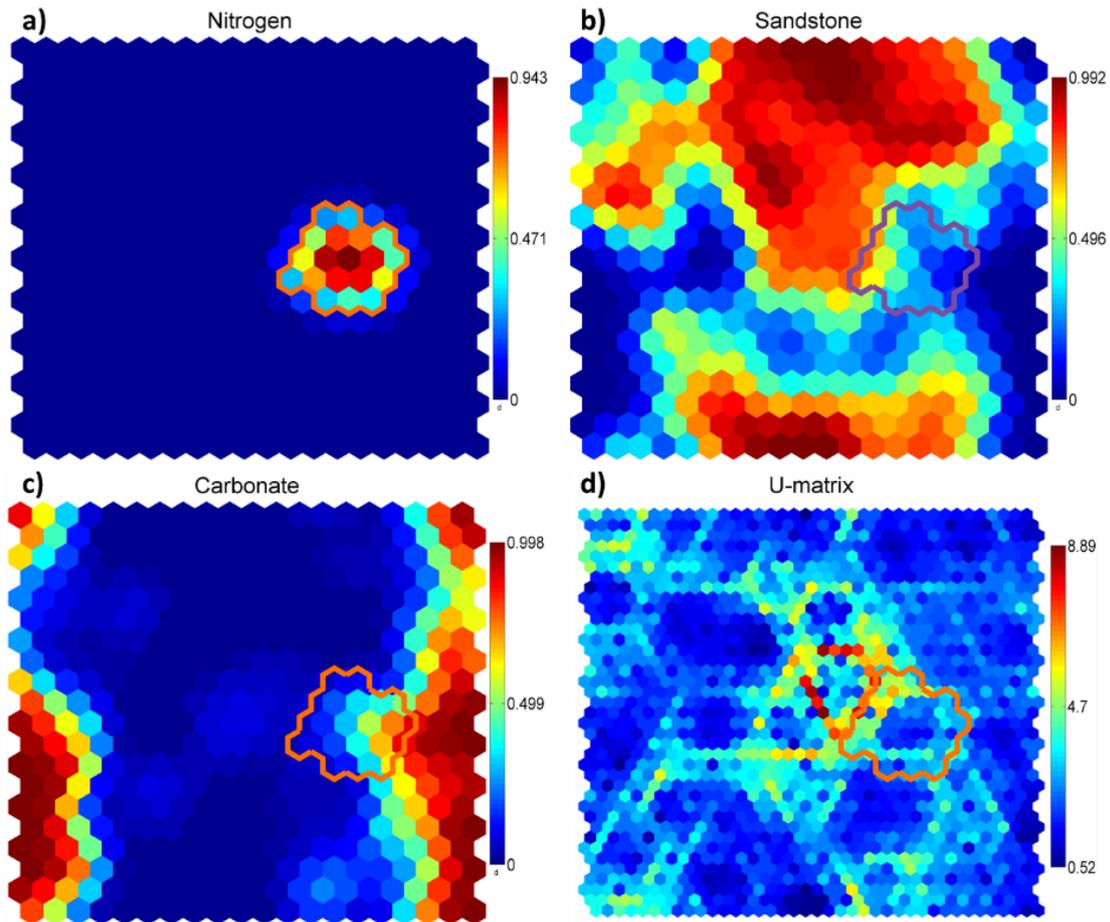


Figura 45. a) *Component plot* referente ao tipo de gás injetado nitrogênio; b) e c) *component plots* referentes às litologias dominantes dos reservatórios; d) Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de gás nitrogênio.

Conforme a Figura 46, dentre os métodos de injeção de gás, N2 apresenta maior média de data de início dos projetos. Também apresenta a menor média de densidade de poços injetores, com número de poços produtores e injetores baixos e na faixa de 4 a 50 poços, com exceção de dois outliers. Esses dois outliers são os campos *onshore* de Hawkins, no Texas, e Elk Hills, na Califórnia, que apresentam grande número de poços produtores e injetores e que possuem,

respectivamente, profundidades baixas da ordem de 4600 e 3000 ft e que, dentre outros fatores, possibilita padrões de injeção de N₂ com menor espaçamento e a um menor custo relativo.

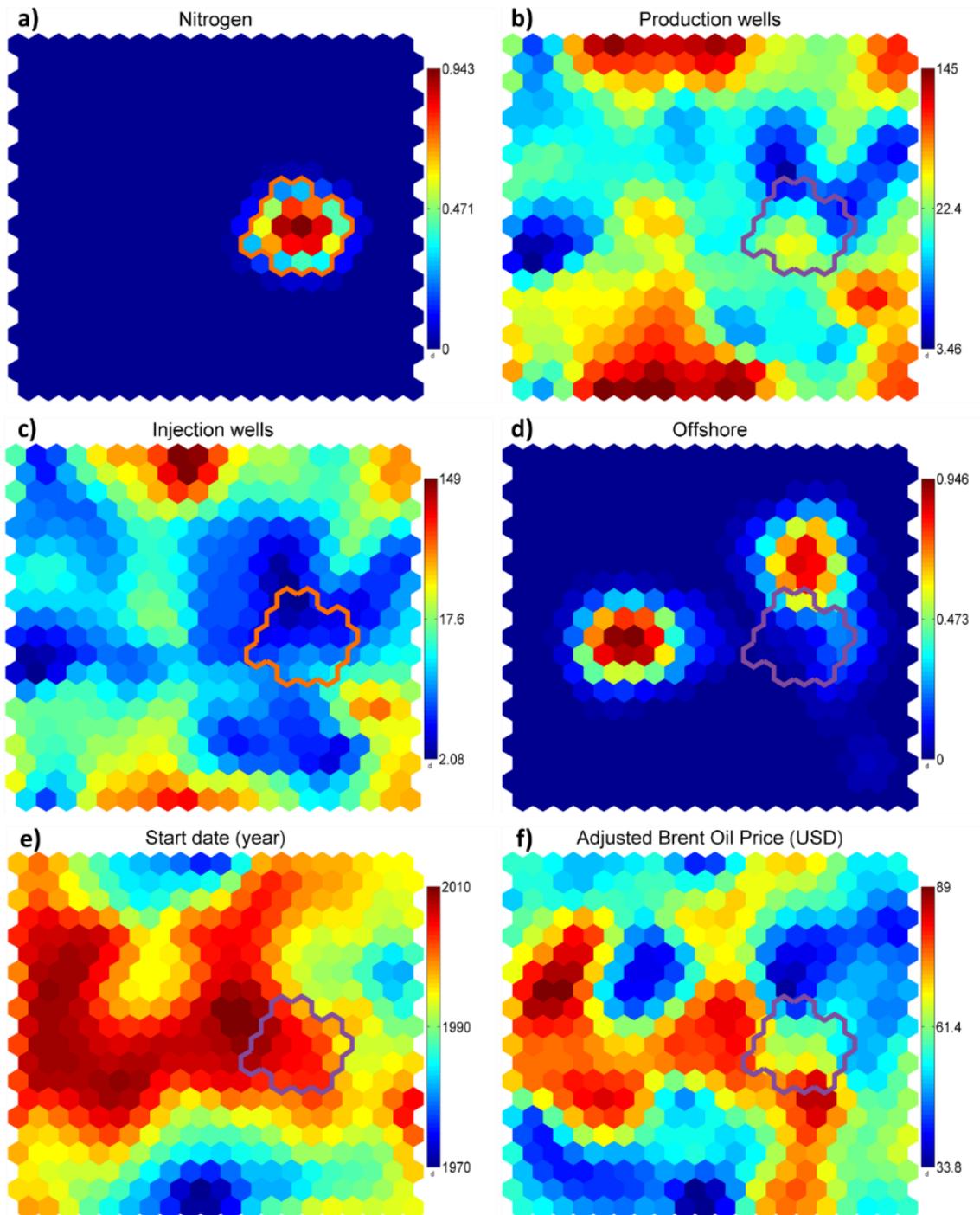


Figura 46. a) *Component plot* referente ao método de gás nitrogênio; b) e c) número de poços produtores e injetores; d) se o projeto é *offshore*; e) ano de início do projeto de EOR; f) o preço do barril ajustado pela inflação no momento de início dos projetos. A linha em destaque delimita os neurônios ativados pelo

método de injeção de gás nitrogênio.

Com base nas *component plots* da Figura 46, os neurônios associados a injeção de N₂ têm 52 USD como preço médio ajustado do barril na data de início de cada projeto. Dentre os primeiros projetos em escala de campo de injeção de N₂ catalogado destaca-se o de Hawkins, com início em 1987, e que teve injeção imiscível de N₂ com o intuito de promover deslocamento de óleo não varrido previamente pela água. Desde então, houveram outras iniciativas, sendo executada por majors (IOCs ou NOCs) e integradas na cadeia de valor em outros locais do mundo. Dentre os projetos *offshore* de injeção de N₂ em escala de campo se destaca o do campo de Cantarell-Akal, no México, que possui porosidade da ordem de 10%, permeabilidade de 2 mD, profundidade de 7500 ft, formação produtora com aproximadamente 4000 ft de espessura, °API 22, se localizando a 80 km da costa.

5.2.3. Injeção de produtos químicos

5.2.3.1. Polímero

Projetos de injeção de polímeros representam 10% da base de dados, com ao todo 168 projetos catalogados, em 16 países. A maior parte dos projetos encontram-se na China, enquanto outros países como EUA, Brasil, países do Mar do Norte, Argentina, Índia, Canadá, Venezuela, Colômbia, Suriname, Angola possuem menor número de projetos.

Com base nas *component plots* da Figura 47, pode-se verificar que os valores médios para as variáveis dos conjuntos de *clusters* com maior quantidade de projetos de injeção de polímeros possuem ranges de variação significativamente menores do que os apresentados nos *boxplots* da Figura 20. Mais de 90% dos BMUs e projetos associados a injeção de polímero possuem valores de porosidade de 17 a 32%, permeabilidades de 90 a 1700 mD, profundidade de 2000 a 7000 ft, temperatura de 65 a 175°F, °API de 16 a 33, viscosidade de 5 a 700 cP, valores estimados altos de garganta de poro, entre 3 e 5,5, e de transmissibilidade estimada, acima de 9000 mD.ft.cP⁻¹.

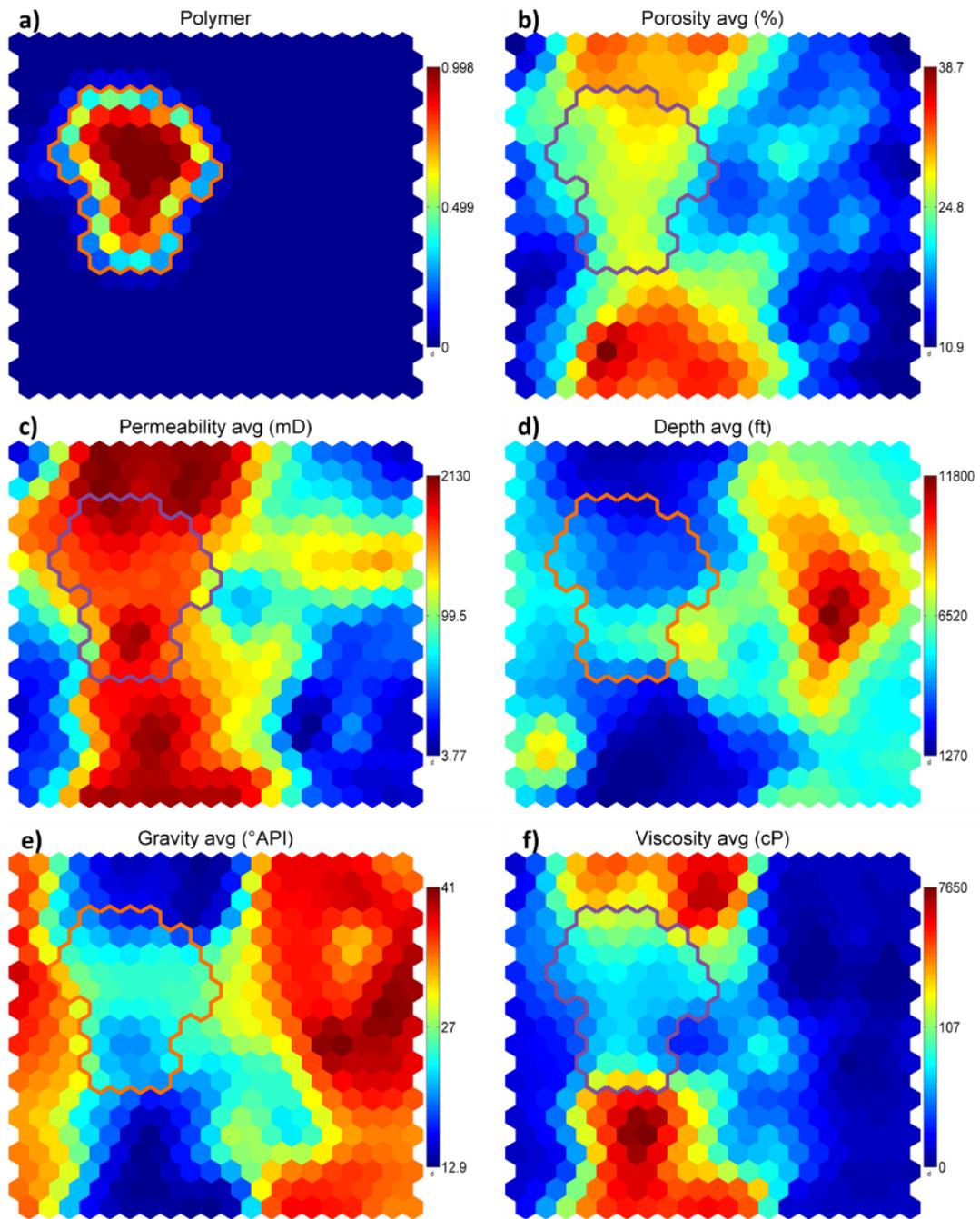


Figura 47. a) *Component plot* referente ao método de injeção de polímero; b) a f) *component plots* de variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero.

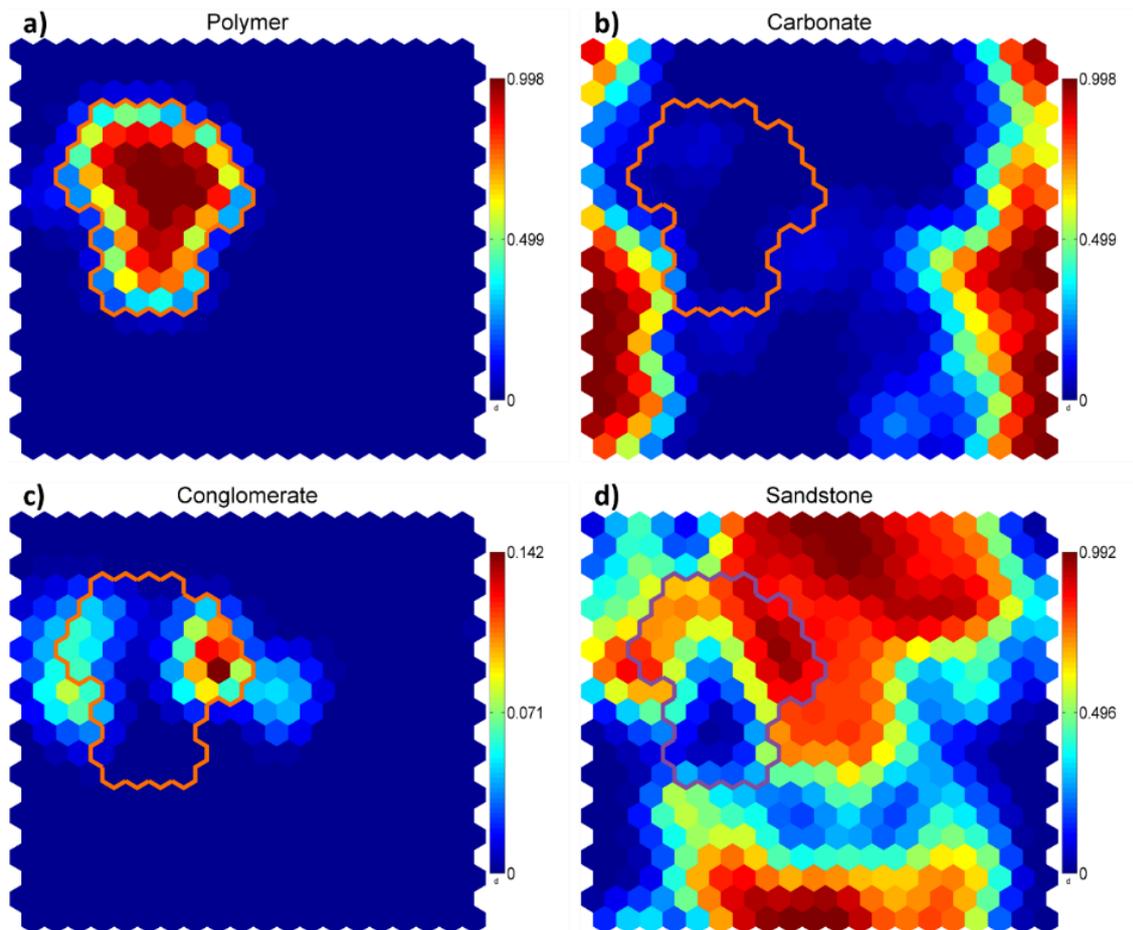


Figura 48. a) *Component plot* referente a injeção de polímero; b) a d) *component plots* referentes às litologias dominantes dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero.

Conforme sintetizam as *component plots* da Figura 48, a injeção de polímero em sua grande maioria é aplicada em reservatórios areníticos; em menor proporção também há aplicações em conglomerados. Há poucos projetos de injeção de polímero em carbonatos, em que esses se caracterizam por apresentarem baixas permeabilidades da matriz e presença de fraturas naturais (ex: campos de Leveland, Mabee). A injeção de polímero geralmente não ocorre em baixas permeabilidades, com exceção de reservatórios carbonáticos: dos 12 projetos em permeabilidades menores do que 100 mD, 9 foram em carbonatos.

Com base nos *boxplots* da Figuras 49, dentre os métodos de EOR químicos, a injeção de polímero apresenta, em média, os maiores valores de transmissibilidade e de garganta de poro estimada. Dentre os métodos químicos, a injeção de polímero possui também os maiores valores médios de garganta de

poro estimada, área de implementação e de números de poços produtores e injetores, e menor média de concentração salina (TDS). Em relação a TDS, há outliers com TDS elevados da ordem de 20000, sendo estes projetos recentes e pilotos que tiveram início de sua injeção posteriormente a 2016; contudo, a maioria dos projetos demonstram ter TDS inferior a 35000 ppm.

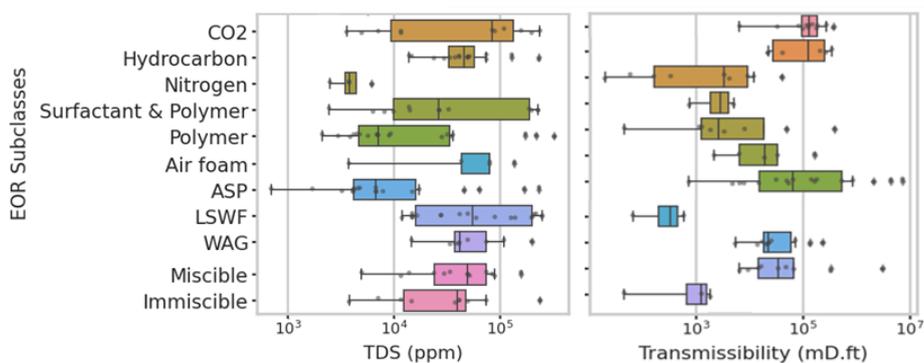


Figura 49. *Boxplots* para as variáveis a) TDS e b) transmissibilidade de acordo categorizados quanto a métodos de EOR.

Injeção de polímero é o método químico com o maior número de projetos *offshore* aplicados em escala de campo, totalizando mais de 20 campos. Observando nas *component plots* da Figura 50, nos neurônios que correspondem à intersecção entre projetos de injeção de polímero e projetos *offshore* destaca-se que essa região coincide com média do preço do barril ajustado pela inflação em torno de 77 USD e que possuem valores médios de porosidade e permeabilidade também acima dos projetos *onshore*. Dentre exemplos de projetos *offshore* de injeção de polímero, estão campos localizados na bacia de Bohai Bay, na China, Dos Cuadras, Mar do Norte e na Bacia de Campos, no Brasil.

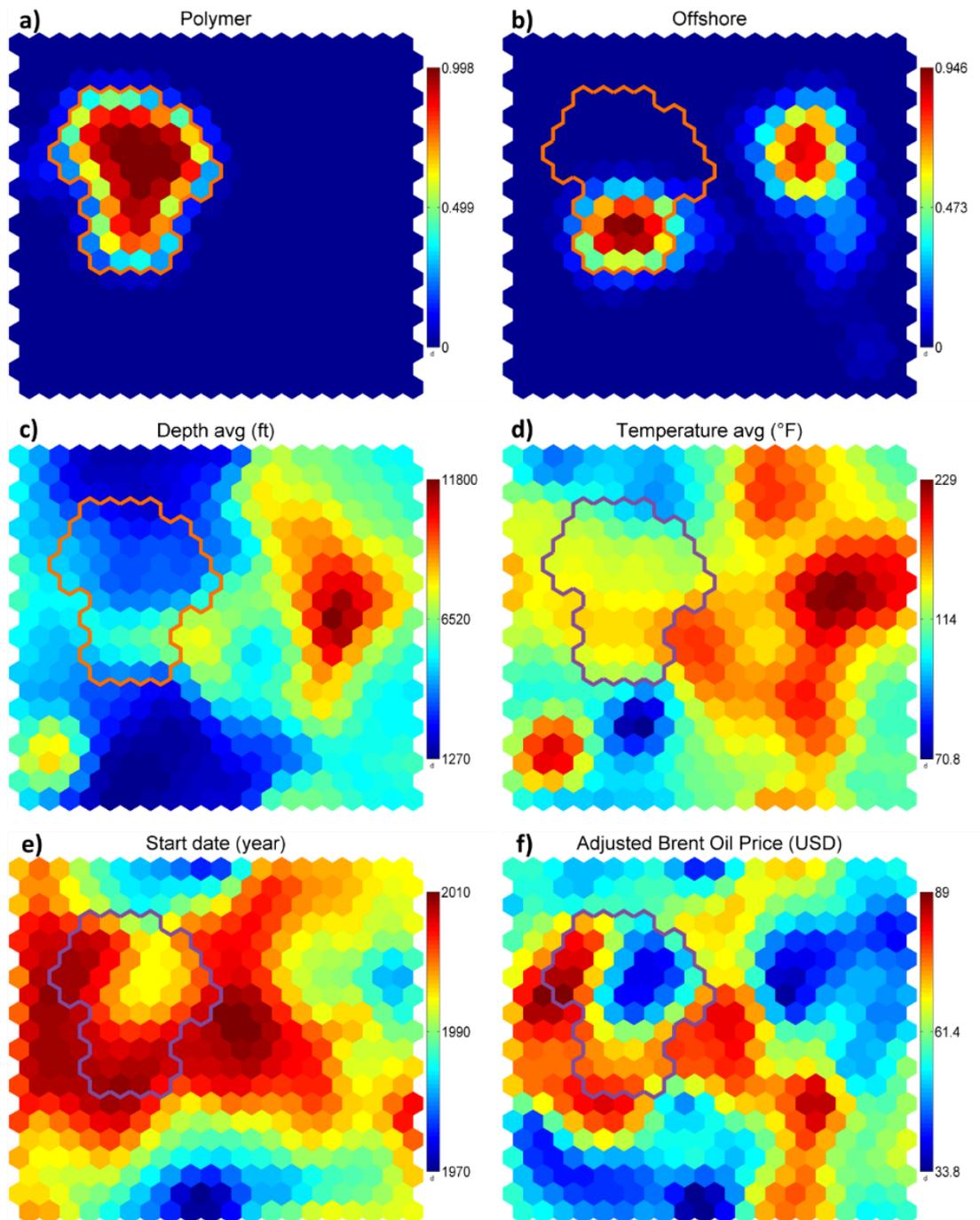


Figura 50. a) *component plot* referente ao tipo de produto químico injetado polímero; b) se o projeto é *offshore*; c) a e) à profundidade e temperatura dos respectivos reservatórios e suas datas inícios de execução; f) preços do barril ajustados pela inflação no início do projeto. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero.

O fator de recuperação incremental associado aos projetos de injeção de polímero se mostra na faixa dos 3 a 10%. Na maioria dos projetos concentrações e volumes pequenos de polímero foram utilizados, permanecendo na escala de pilotos. Além disso, pequenas alterações e melhorias são difíceis de se identificar

e atribuir como resposta da injeção de polímero. Contudo, há casos de campos com claras respostas bem sucedidas, como por exemplo Marmul, Oerrel, Courtenay, Daqing. Tais campos apresentam características comuns, como médias de permeabilidades da ordem de 800 mD, temperaturas do reservatórios baixas de até 140°F, altas saturações de óleo no início do projeto, baixa salinidade (TDS), obtendo fatores de recuperação incremental de 11 até 30%.

É possível visualizar nas *component plots* da Figura 50 também que há uma associação dos projetos de injeção de polímero de datas de início mais recentes e preços médios a altos do barril ajustado pela inflação com maiores valores profundidades e temperaturas de reservatório. Nesse conjunto de reservatórios, os *clusters* alcançam profundidades médias de 6000 ft, variando de 5000 a 7000 ft, e temperaturas médias de até 160°F, havendo também reservatórios dentro desses *clusters* que se tratam de projetos *offshore* e/ou com profundidades e temperaturas maiores, além de permeabilidades e transmissibilidades menores quando se comparados ao restante dos projetos de injeção de polímero. Essa tendência de aplicação da injeção de polímero em profundidades maiores à medida que o tempo avança e/ou preço do barril se eleva, pode também ser observada nos scatterplots da Figura 51 referente aos métodos químicos, em que $\rho=0.408$ para preço ajustado do barril vs profundidade e $\rho=0.336$ para data de início do projeto vs profundidade; para a variável temperatura essas correlações são de $\rho=0.116$ e 0.140 , respectivamente. Altas significâncias e correlação para tais conjuntos de variáveis não são obtidas para outros métodos de EOR, com por exemplo para os métodos de injeção de gás: $\rho=0.032$ para preço do barril vs profundidade e $\rho=0.011$, além de baixas significâncias estatísticas. Tal fato pode indicar o desafio tecnológico e econômico de se aplicar polímero em profundidades e temperaturas maiores, e um ganho de maturidade e avanço da tecnologia, e cuja atratividade econômica e aplicação nesses contextos de elevadas profundidades e temperaturas tende a ser mais alta e estimulada em cenários de preços do barril mais altos.

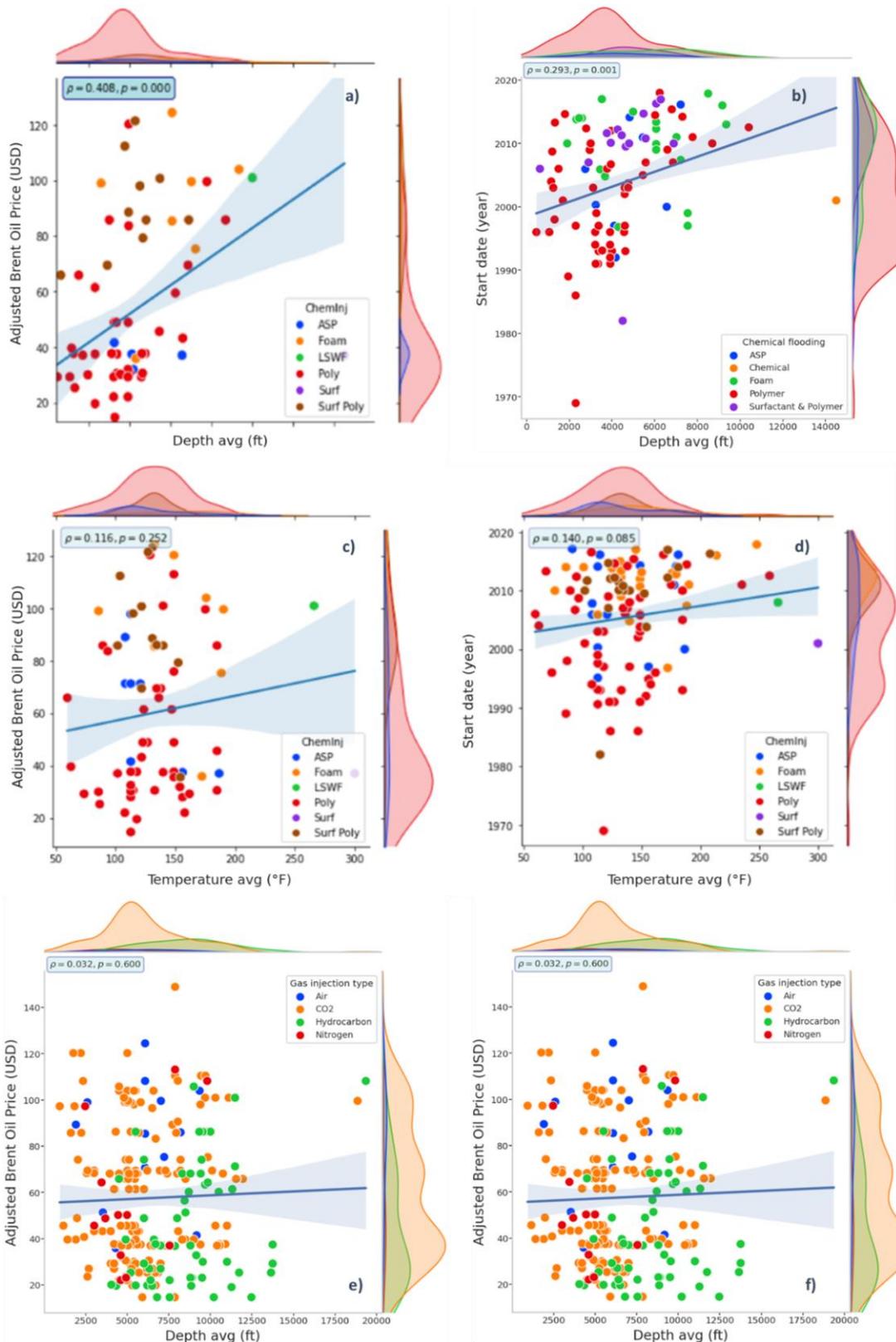


Figura 51. Para os dados de projetos que tiveram injeção de métodos químicos, scatterplots da variável profundidade com a) preço ajustado do barril preço ajustado do barril e b) data de início do projeto; da variável temperatura com c) preço ajustado do barril preço ajustado do barril e d) data de início do projeto. Para os dados de projetos que tiveram projetos com métodos de injeção de gás,

scatterplots da variável profundidade com e) preço ajustado do barril preço ajustado do barril e f) data de início do projeto.

A injeção de polímero possui aplicação em campos e pesquisa há mais de 40 anos, conforme pode ser visto nos gráficos de histórico de projetos e de papers nas Figuras 3 e 15. Em especial, nos anos 1990, o número de aplicações sofreu uma queda com a manutenção do preço do barril baixo. A injeção de polímero voltou a ter expansão por volta do ano 2005, podendo ser associado tanto ao aumento do preço do barril, como também ao ganho de maturidade da tecnologia e sua forte aplicação em países como a China.

Nas *component plots* da Figura 52 observa-se que nos neurônios referentes a injeção de polímero e em relação à escala de projeto, tende a haver maior associação com projetos pilotos do que se comparado a projetos em escala de campo. Observa-se que a maior parte dos neurônios associados a injeção de polímero possuem baixos valores próximos de nulos ou próximo de nulos para escala de campo, com alguns *clusters* na parte esquerda inferior tendo valores médios para a escala de campo. Além disso, por inspeção, verificase que 62% desses projetos em escala de campo foram aplicados na China.

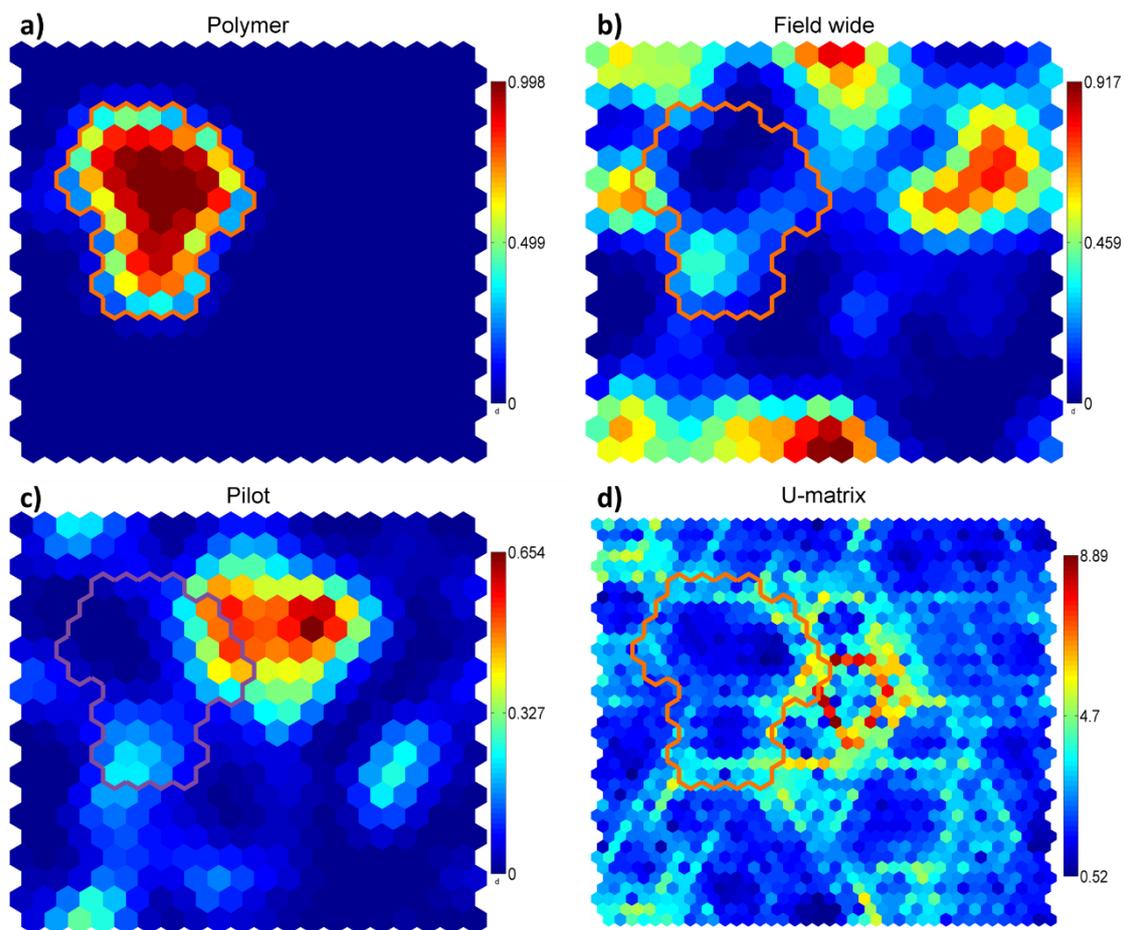


Figura 52. *Component plots* referentes a) ao tipo de químico injetado (polímero) e à escala do projeto - se implementado na forma de b) projeto piloto ou c) em escala de campo); d) Matriz-U. A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero.

Conforme pode-se observar nos mapas da Figura 53, os projetos de injeção de polímero foram realizados preponderantemente por empresas de grande porte, principalmente por NOC e IOC, e em geral também integradas na cadeia de exploração, produção, transporte, refino e distribuição. Há pouca ou nenhuma atuação de empresas pequenas e médias; e apenas em projetos de injeção de polímero com datas de início mais recentes, e preço do barril relativamente maior, é que se verifica maior atuação de empresas focadas no segmento de exploração e produção.

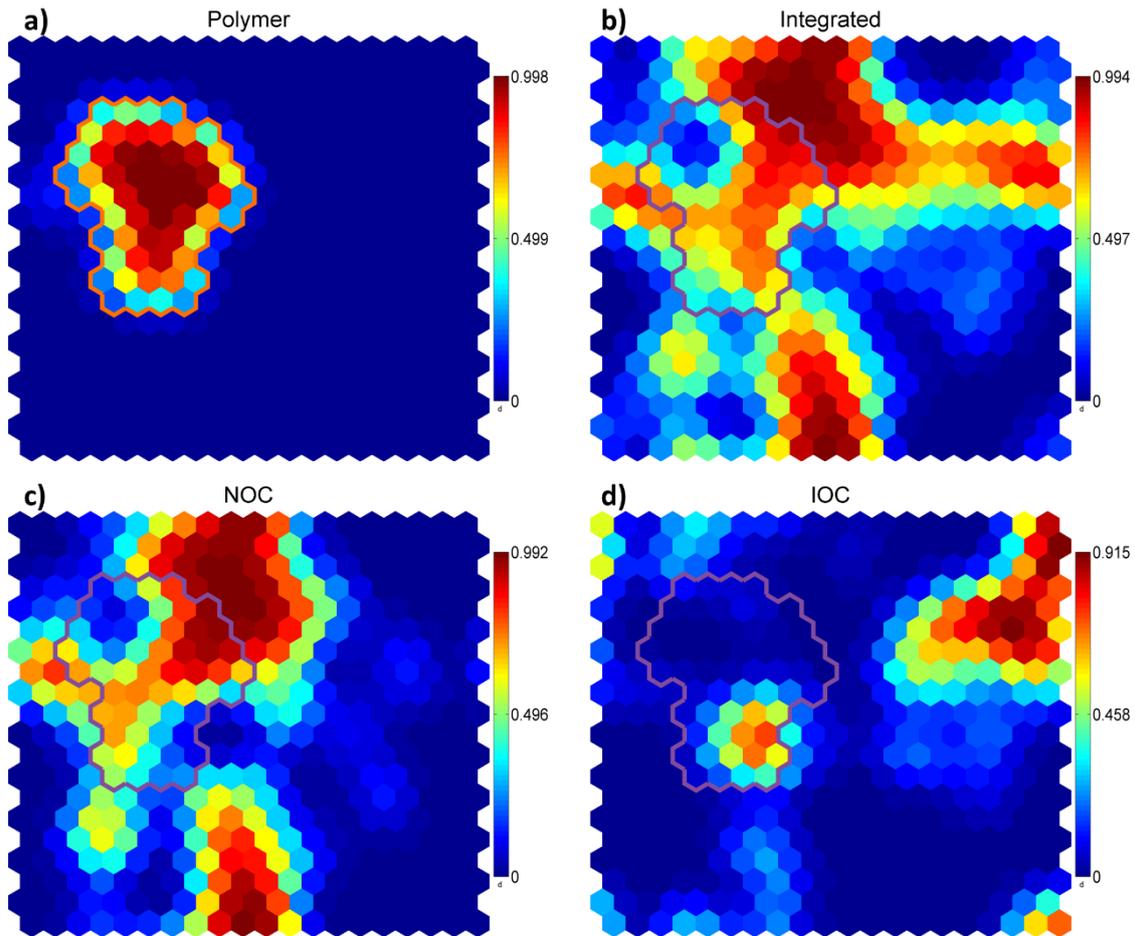


Figura 53. a) *Component plot* referente ao método de injeção de polímero; b) *component plot* referente ao segmento de atuação (se integrada na cadeia de valor); d) e e) tamanho da companhia operadora, se International Oil Company (IOC) ou National Oil Company (NOC). A linha em destaque delimita os neurônios ativados pelo método de injeção de polímero.

O fator de recuperação incremental associado aos projetos de injeção de polímero se mostra na faixa dos 3 a 10%. Na maioria dos projetos concentrações e volumes pequenos de polímero foram utilizados. Além disso, pequenas alterações e melhorias são difíceis de se identificar e atribuir como resposta da injeção de polímero. Contudo, há casos de campos com claras respostas bem sucedidas, como por exemplo Marmul, Oerrel, Courtenay, Daqing. Tais campos apresentam características comuns, como médias de permeabilidades da ordem de 800 mD, temperaturas do reservatórios baixas de até 140°F, altas saturações de óleo no início do projeto, baixa salinidade (TDS), obtendo fatores de recuperação incremental de 11 até 30%.

5.2.3.2. *Surfactante e Polímero (SP)*

Projetos de injeção de Surfactante e Polímero conjuntamente (SP) representam 8% da base de dados, com ao todo 124 projetos catalogados, em 9 países. A maior parte dos projetos encontram-se principalmente na China e EUA, seguidos por Indonésia, UAE, Reino Unido, Hungria e Malásia.

Com base nas *component plots* da Figura 54, pode-se verificar que os valores médios para as variáveis dos conjuntos de *clusters* com maior quantidade de projetos de injeção de polímeros possuem um intervalo de valores pequeno se comparado aos outros métodos de EOR. Mais de 90% dos *clusters* e projetos de injeção de SP possuem valores de porosidade de 11 a 23%, permeabilidades de 10 a 110 mD, profundidade de 3000 a 6000 ft, temperatura de 90 a 140°F, °API de 28 a 38, viscosidade de 1 a 20 cP, valores estimados altos de garganta de poro (entre 2 e 3,5).

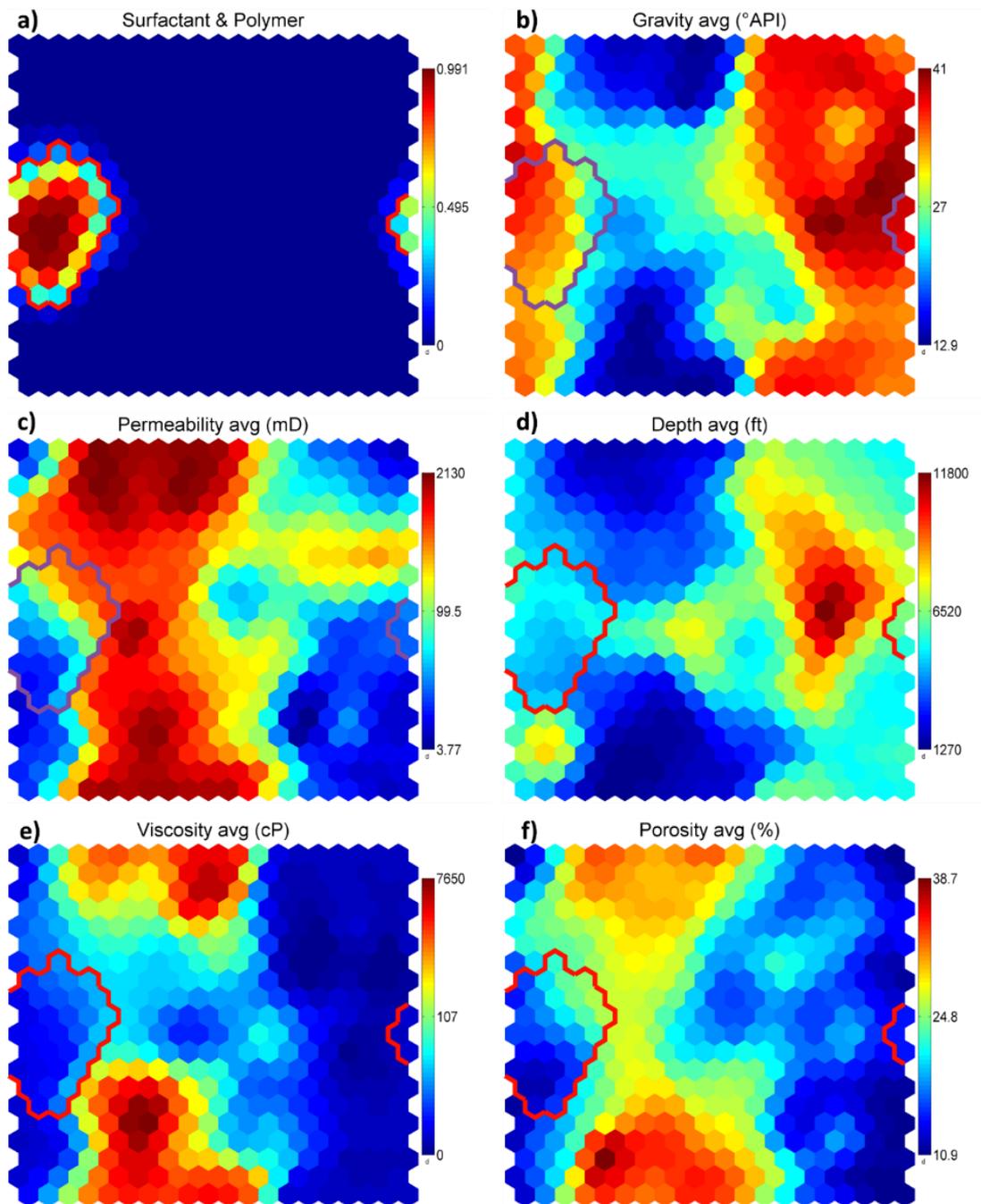


Figura 54. a) *component plot* referente ao método de injeção de surfactante e polímero; b) a f) *component plots* das demais variáveis básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção de surfactante e polímero.

A injeção de SP possui aplicações tanto em reservatórios carbonáticos de como também areníticos. Dentre os métodos químicos, SP é a tecnologia com maior aplicação em reservatórios carbonáticos, tanto em porcentagem quanto em número absolutos. Contudo, dentre os métodos químicos, SP é o com maior porcentagem de aplicação de pilotos, em detrimento da aplicação em escala de

reservatório ou escala comercial. A maior parte dos projetos de SP teve sua aplicação a partir de 2008, com preços corrigidos do barril do petróleo acima de 70 dólares, conforme mostra as *component plots* da Figura 55. A injeção de SP possui apenas 5 projetos catalogados de aplicação *offshore* e que possuem em escala de campo/comercial, dentre os quais dois estão localizados em Bohai Bay, na China, e em campos com \circ API médio igual a 20. Os projetos de injeção de SP em sua maioria tiveram previamente injeção de água, com poucos casos em que se partiu diretamente da produção primária. A maior parte dos projetos de SP foram conduzidos por NOCs, seguido por IOCs.

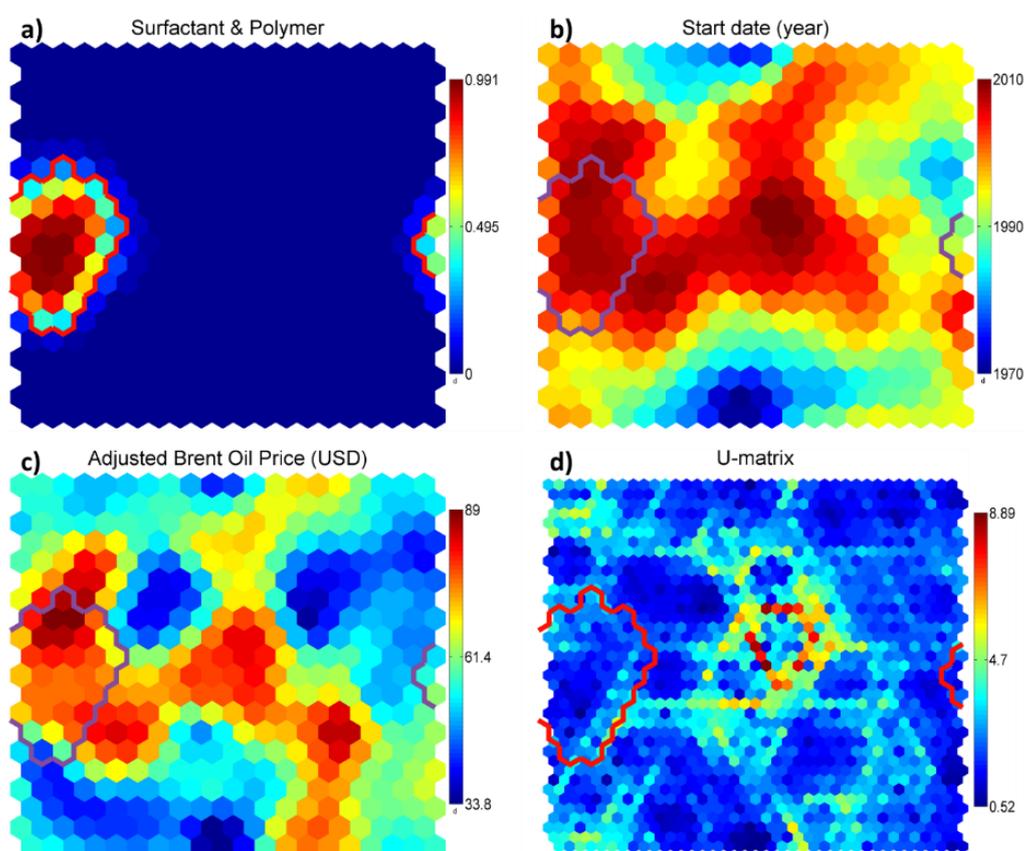


Figura 55. a) *component plots* do tipo de químico injetado (surfactante e polímero), b) e c) data de início e preço ajustado do barril no ano de início dos projetos; d) Matriz-U. A linha em destaque delimita os neurônios ativado pelo método de injeção de surfactant e polímero.

5.2.4. Outros métodos

5.2.4.1. Microbial

Projetos de injeção microbial representam 2% da base de dados, com ao todo 30 projetos catalogados, em 4 países. A maior parte dos projetos encontram-se principalmente na China e nos EUA, com aplicações em menor número na Rússia e Brasil.

Com base nas *component plots* da Figura 56, pode-se verificar que os valores médios para as variáveis dos conjuntos de neurônios com maior quantidade de projetos de injeção microbial possuem um intervalo de valores menor do que em relação aos *boxplots*. Mais de 90% dos neurônios e projetos de injeção de polímero possuem valores de porosidade de 14 a 21, permeabilidades de 90 a 700 mD, profundidade de 2000 a 6000 ft, temperatura de 85 a 180°F, °API de 25 a 31, viscosidade de 5 a 200 cP, valores estimados de garganta de poro entre 2.8 e 4 e de mobilidade do óleo de 1 a 20.

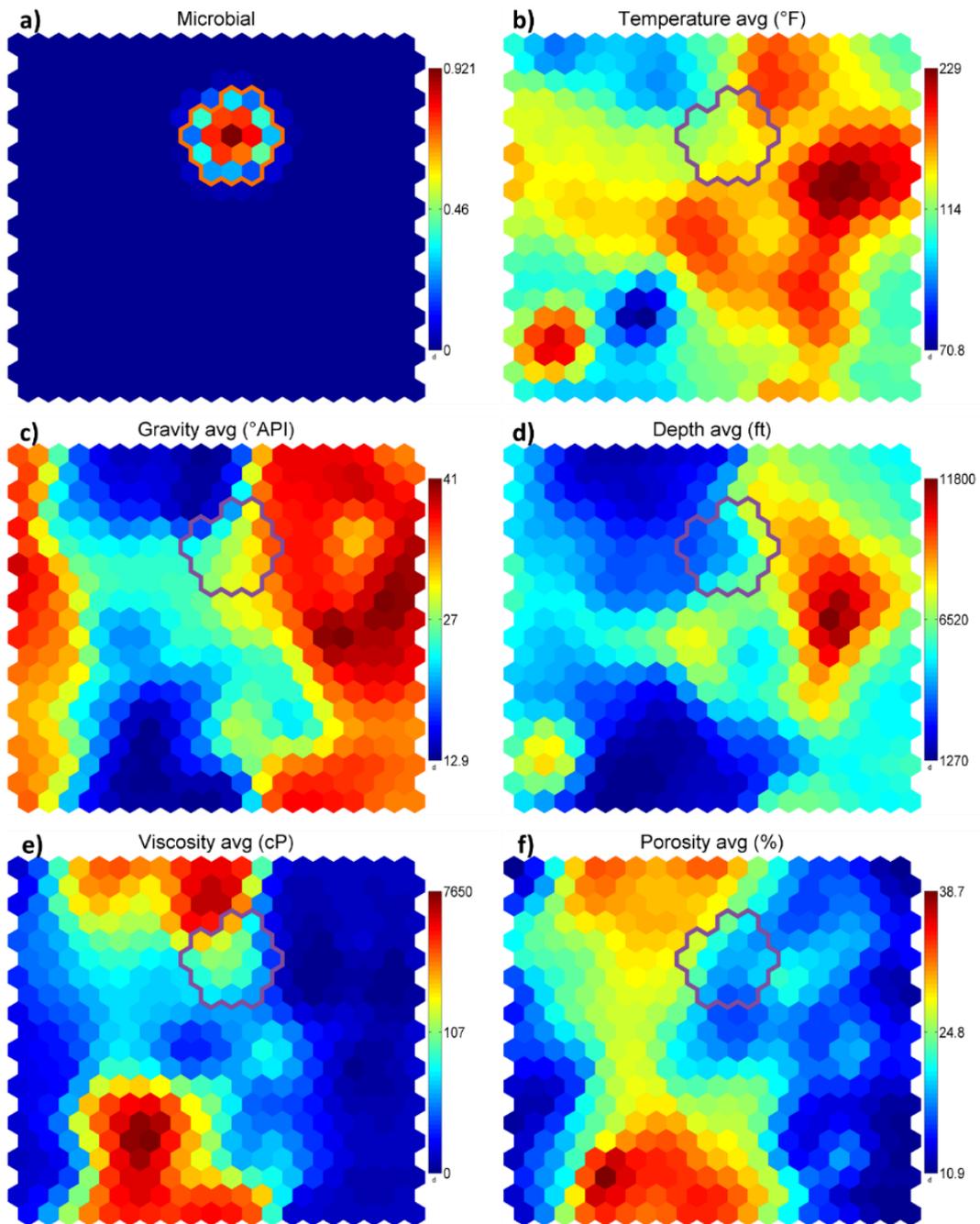


Figura 56. *Component plots* referente a a) injeção microbial e b) a f) variáveis de propriedades básicas dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção microbial.

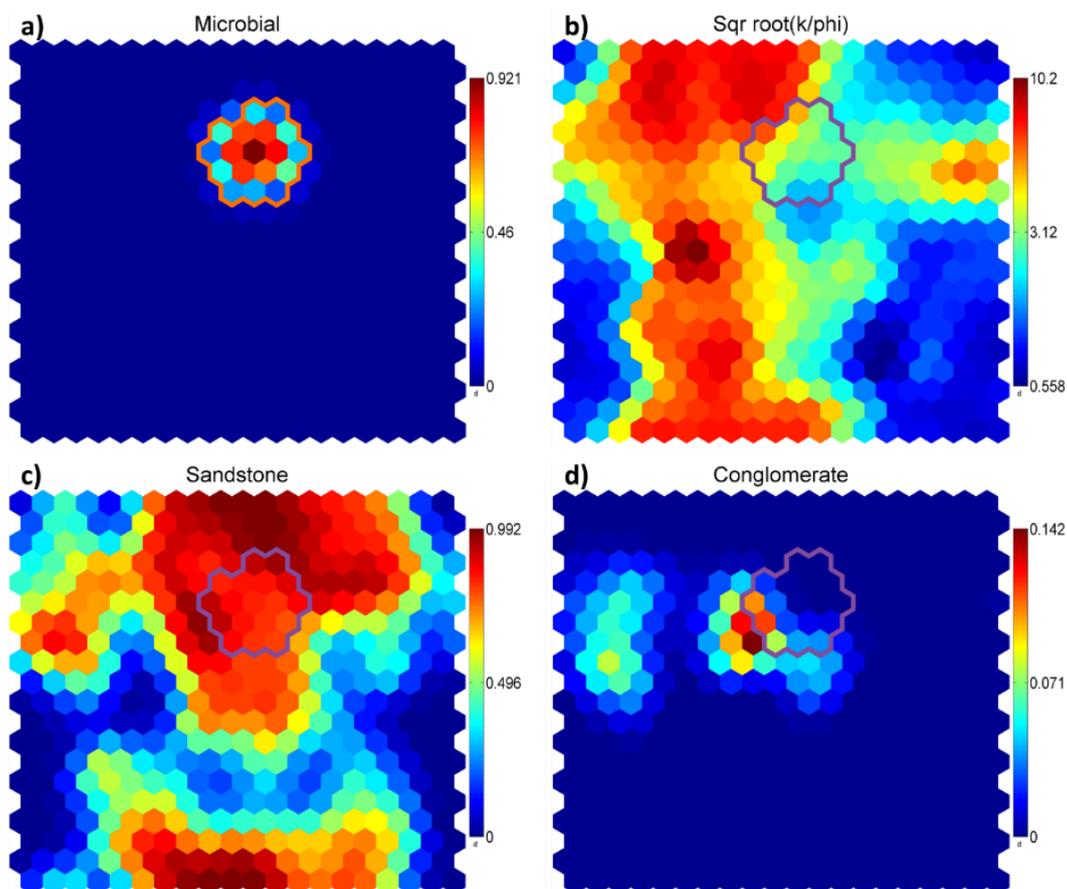


Figura 57. *Component plots* referente a a) injeção microbiana e b) a d) variáveis referentes às litologias dominantes dos reservatórios. A linha em destaque delimita os neurônios ativados pelo método de injeção microbiana.

Conforme sintetiza as *component plots* da Figura 57, o método microbiano em sua grande maioria é aplicada em reservatórios areníticos, molháveis à água; em menor proporção também há aplicações em arenitos conglomerados. Conforme Figura 56, é possível observar que a injeção microbiana, se comparada aos demais métodos de EOR classificados como “Outros”, apresenta limitação na sua profundidade e na temperatura dos reservatórios em que são aplicados. Os neurônios com maior influência de projetos relacionados a injeção microbiana possuem baixos valores para a variável *offshore*, em decorrência de a maior parte dos projetos microbianos serem em ambiente *onshore*. Dentre os poucos projetos *offshore*, destacam-se as aplicações nos campos de Norne, no mar do Norte com bactérias redutoras de nitrato, e Bokor, na Malásia com o uso de recuperação cíclica das bactérias injetadas (AWAN, 2008; KOWALEWSKI et al., 2006; PATEL et al., 2015; RASSENFOSS, 2011).

É possível visualizar nas *component plots* da Figura 58 também que os neurônios com maior influência de projetos microbiais relacionam-se a projetos de injeção microbiana com datas de início variando dos anos 1990 até a atualidade. Tais neurônios também mostram a associação de métodos microbiais principalmente à escala de projetos pilotos, e valores ou nulos para a variável de escala de campo. Dentre os poucos projetos implementados em escala de campo, destaca-se o de Baolige, na China (KE et al., 2018), de óleo pesado com viscosidade média de 157 cP, porosidade de 18% e temperatura da formação de 122°F, em que bactérias exógenas e nutrientes foram injetados através de 78 poços de injeção durante período de 40 a 60 dias, seguido por injeção de água; tais condições se mostraram favoráveis à aplicação microbiana, e o projeto obteve uma produção extra acumulada de petróleo da ordem de 210000 toneladas no período de 43 meses. Conforme pode-se observar nos mapas da Figura 58, os projetos de injeção microbiana foram realizados preponderantemente por empresas de grande porte, como NOCs, a partir principalmente dos anos 90 na China.

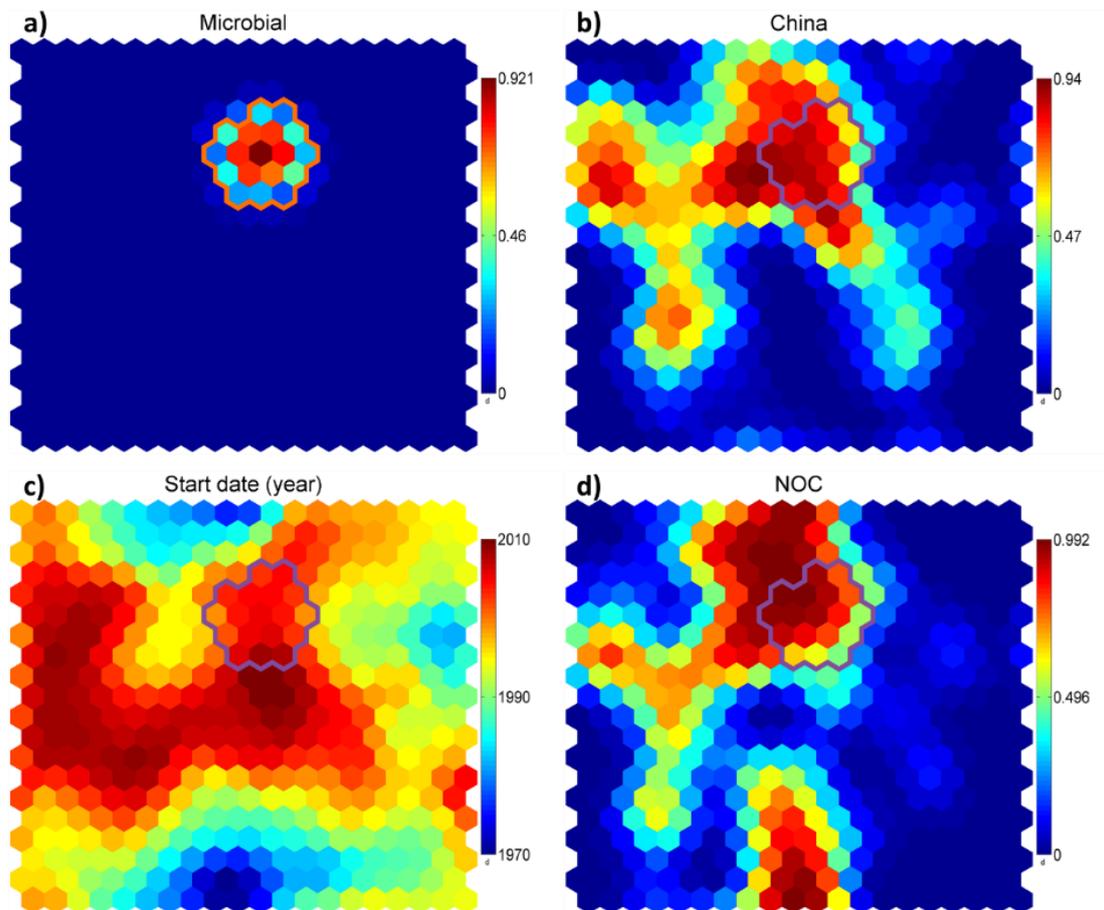


Figura 58. *Component plots* referente a a) injeção microbiana; b) implementação de projetos na China, c) ano de início dos projetos; d) se a empresa operadora de cada reservatório é classificada como National Oil Company.

5.2.4.2. Água de baixa salinidade (LSWF)

Projetos de LSWF representam 2% da base de dados, com ao todo 33 projetos catalogados, em 9 países, conforme mostra as figuras 14 e 16. A maior parte dos projetos encontram-se nos EUA e em países do Mar do Norte e Oriente Médio.

Com base nas *component plots* da Figura 59, pode-se verificar que os neurônios que representam projetos de injeção de polímero possuem valores de porosidade de 14 a 25%, permeabilidades de 50 a 1000 mD, profundidade de 4000 a 8000 ft, temperatura de 90 a 220°F, °API de 25 a 32 e viscosidade de 0.5 a 40 cP.

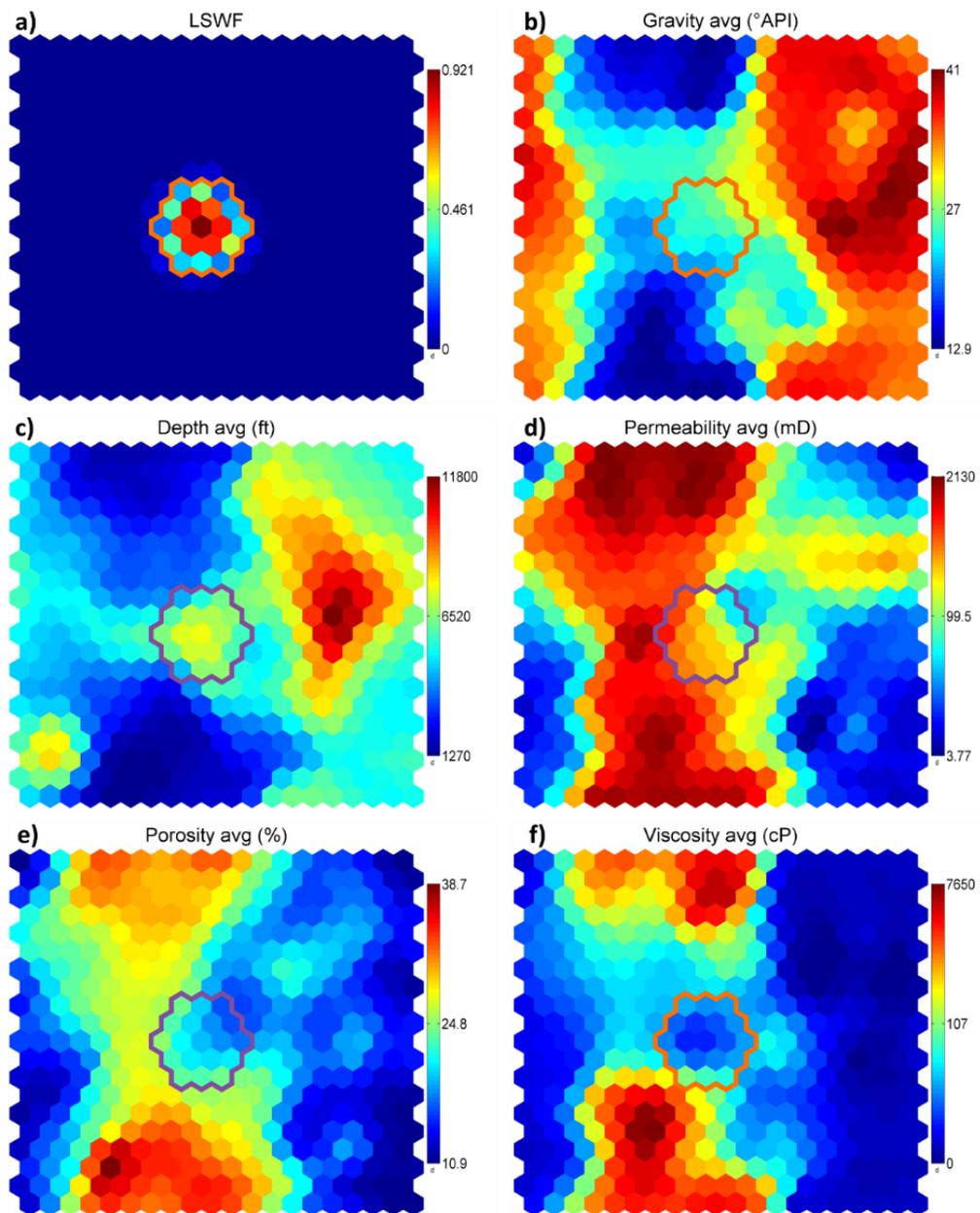


Figura 59. *Component plots* referentes a a) método de injeção de água de baixa salinidade (LSWF); b) a f) propriedades básicas dos reservatórios.

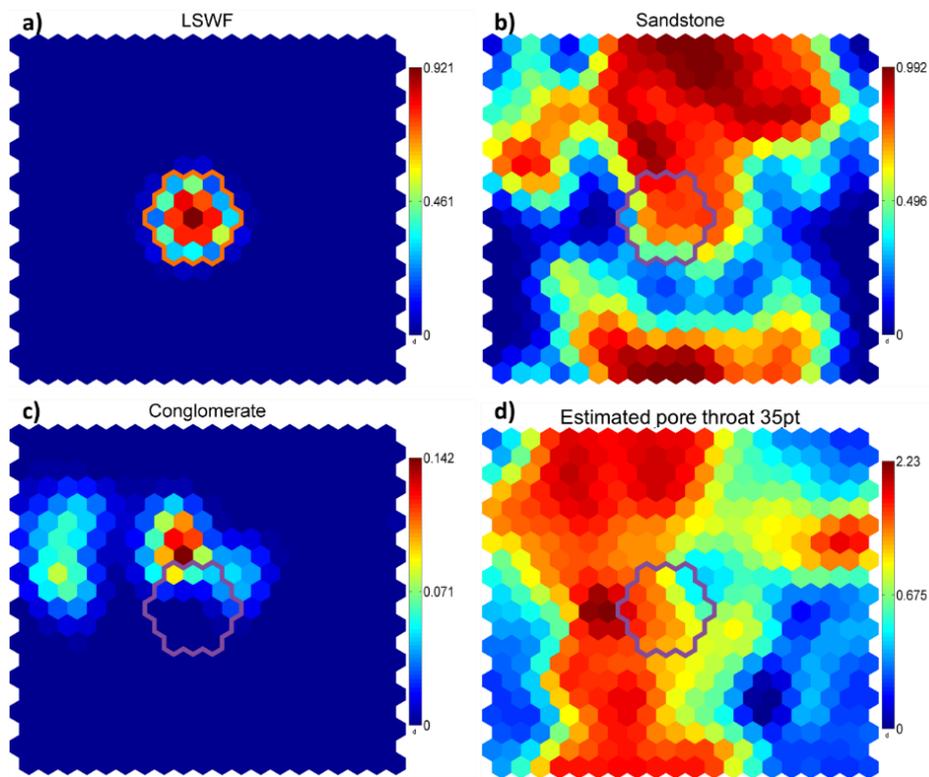


Figura 60. *Component plots* referentes a a) método de injeção de água de baixa salinidade (LSWF); b) a c) litologias dominantes dos reservatórios e d) tamanho estimado da garganta de poro.

Conforme as *component plots* da Figura 60, os neurônios que representam a aplicação de projetos de LSWF se mostram associados à aplicação predominantemente em reservatórios areníticos, com pequena presença de aplicações em carbonatos. Já de acordo com a Figura 61, o emprego de LSWF em escala de campo teve seu início a partir dos anos 2000, com crescimento na sua utilização a partir de 2010, período esse com o preço do barril elevado. Apesar de ser recente a sua aplicação, LSWF destaca-se por já apresentar aplicações de campo em ambiente *offshore*, como os campos areníticos de Endicott no Alasca e Snorre, Heidrun e Clair Ridge no Mar do Norte (CHEN et al., 2021). Os campos com projetos de LSWF foram em sua maioria conduzidos por IOCs e em menor número por NOCs, geralmente com fonte próxima de captação de água (como água do mar), e com nenhuma aplicação catalogada realizada por operadoras independentes de pequeno e médio porte.

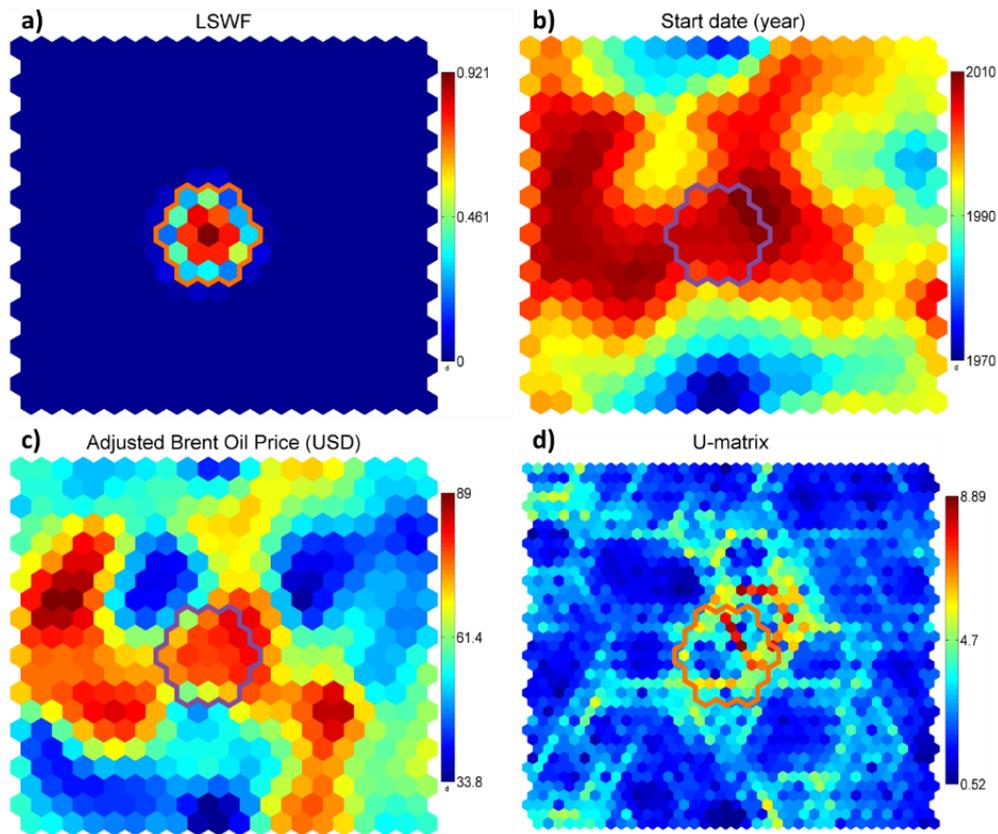


Figura 61. *Component plots* referentes a a) método de injeção de água de baixa salinidade (LSWF); b) data de início do projeto; c) preço do barril no início do projeto ajustado pela inflação; d) Matriz-U.

5.3. Análises derivadas da aplicação de similaridade do cosseno e clusterização hierárquica na base de projetos de EOR

5.3.1. *Clusterização de países e empresas aplicadoras de EOR*

A clusterização dos países com base nas características dos projetos de EOR implementados em cada localidade permitiu o agrupamento e sua visualização na forma de dendrogramas, como mostra a Figura 62. Países próximos na visualização do dendrograma, desde que no interior da mesma árvore, indicam maiores similaridades entre si em relação às características dos projetos de EOR de cada uma dessas nações, como por exemplo Líbia e UAE, e Canadá e EUA. Além disso, quanto menor a altura de um grupo de países em relação a um agrupamento vizinho, maiores as similaridades encontradas nos projetos de EOR dos constituintes desses grupos. Duplas de países como Reino Unido e

Noruega, Rússia e Abu Dhabi, Síria e Arábia Saudita, e Canadá e EUA possuem alturas baixas dos ramos, indicando uma elevada similaridade dos países de cada dupla. Por outro lado, duplas de países também vizinhos no dendrograma como México e Iraque, e Indonésia e Alemanha possuem maior altura do ramo, o que indica menores similaridades entre os países dessas duplas, e reduzindo a possibilidade de inferências de similaridade em afirmações a respeito de tais casos.

É possível observar a formação de agrupamentos de países com características similares em relação aos projetos de EOR implementados. Dentre esses grupos, identificou-se cinco conjuntos de países nos quais se verifica similaridades significativas de contexto, a saber:

- a) Canadá e EUA, países com histórico longo de produção e grande quantidade de reservatórios depletados; infraestrutura consolidada e disponibilidade de fluido injetante; fortes incentivos governamentais para a aplicação da década de EOR, especialmente na década de 1990; atuação de espectro variado de empresas e com expertise; intensa aplicação de EOR, tanto de métodos térmicos, químicos e injeção de gás;
- b) Reino Unido, Noruega e Escócia, que possuem campos produtores no Mar do Norte; presença de regulação com incentivos fiscais e aplicação de EOR predominantemente *offshore* e de injeção de gás obtido da produção de campos vizinhos;
- c) Trindade e Tobago, China, Colômbia e Índia, que apresentam aplicação de EOR predominantemente em campos com óleo médio a pesado de métodos térmicos e químicos;
- d) Venezuela e Brasil, países que apresentaram número significativo de aplicações em escala de campo de métodos térmicos, com mais recentemente injeção de gás e projetos pilotos de produtos químicos; contudo, é importante notar a elevada dissimilaridade no dendrograma entre as duas nações;
- e) Conjunto de países que possuem amplas reservas de petróleo, que tendem a apresentar baixos custos de extração (*lifting costs*), e menores fatores de recuperação; em geral, apresentam aplicações apenas mais recentemente e pontuais de EOR;

Tais grupos e embasamento em literatura dessas classificações fornecidas pelo dendrograma são avaliados em maior profundidade na sessão de Discussões.

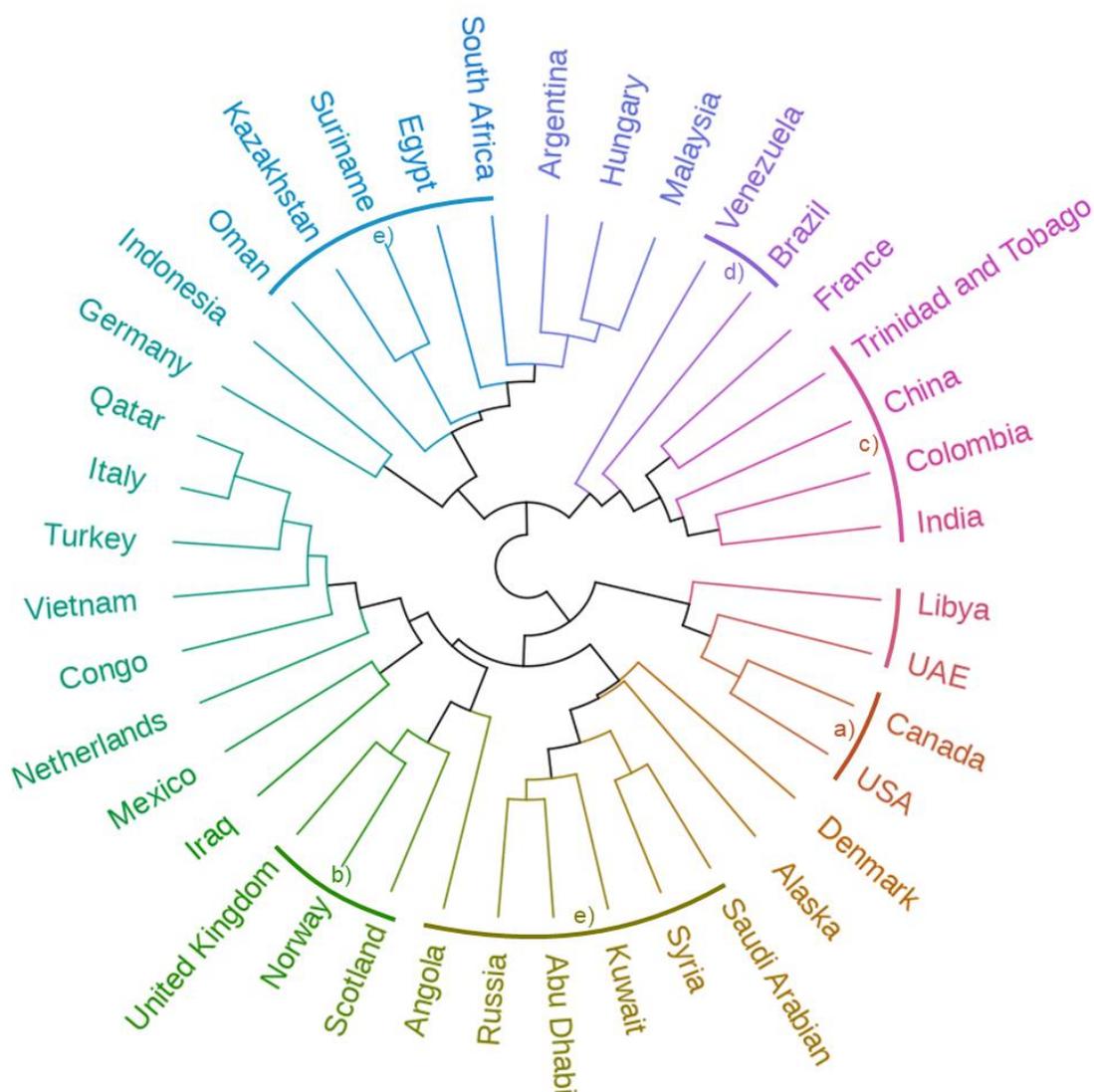


Figura 62. Clusterização dos países identificados com projetos de EOR e representação na forma de dendrogramas, com a identificação de cinco grupos principais de países.

A clusterização das empresas com base nas características dos projetos de EOR implementados por cada uma delas permitiu o agrupamento e sua visualização na forma de dendrogramas, conforme a Figura 63. Empresas próximas na visualização do dendrograma, desde que no interior da mesma árvore, e com menor altura do ramo uma em relação à outra, apresentam maiores similaridades entre si em relação às características dos projetos de EOR de cada uma dessas

empresas, como por exemplo: Breitburn Energy, e Merit Energy e Denbury Resource. Já duplas de empresas com maiores alturas dos ramos uma em relação à outra, as similaridades tendem a ser menores, como no caso de Chevron e Shell, India Oil and Natural Gas Corporation e Petrobras, British Petroleum e Exxon Mobil, Equinor e Total. Apesar de constituírem grupos de empresas com portfolio e estratégias competitivas semelhantes, possuem grande dispersão global e variedade em relação aos ativos que possuem e aos métodos de EOR que foram empregados. É possível observar a formação de agrupamentos de empresas com características similares em relação aos projetos de EOR implementados. Dentre esses grupos, identificou-se quatro principais conjuntos de empresas nos quais se verifica similaridades significativas de contexto, a saber: a) empresas de grande porte, NOCs e IOCs, geralmente integradas na cadeia de valor; b) empresas de grande porte, porém em sua maior parte independentes e focadas no segmento de exploração e produção; c) empresas de porte médio a grande, geralmente com atuação local; d) empresas próximas lateralmente no dendrograma mas que apresentam maior nível de dissimilaridade entre si se comparado com as empresas dos outros três grupos anteriores.

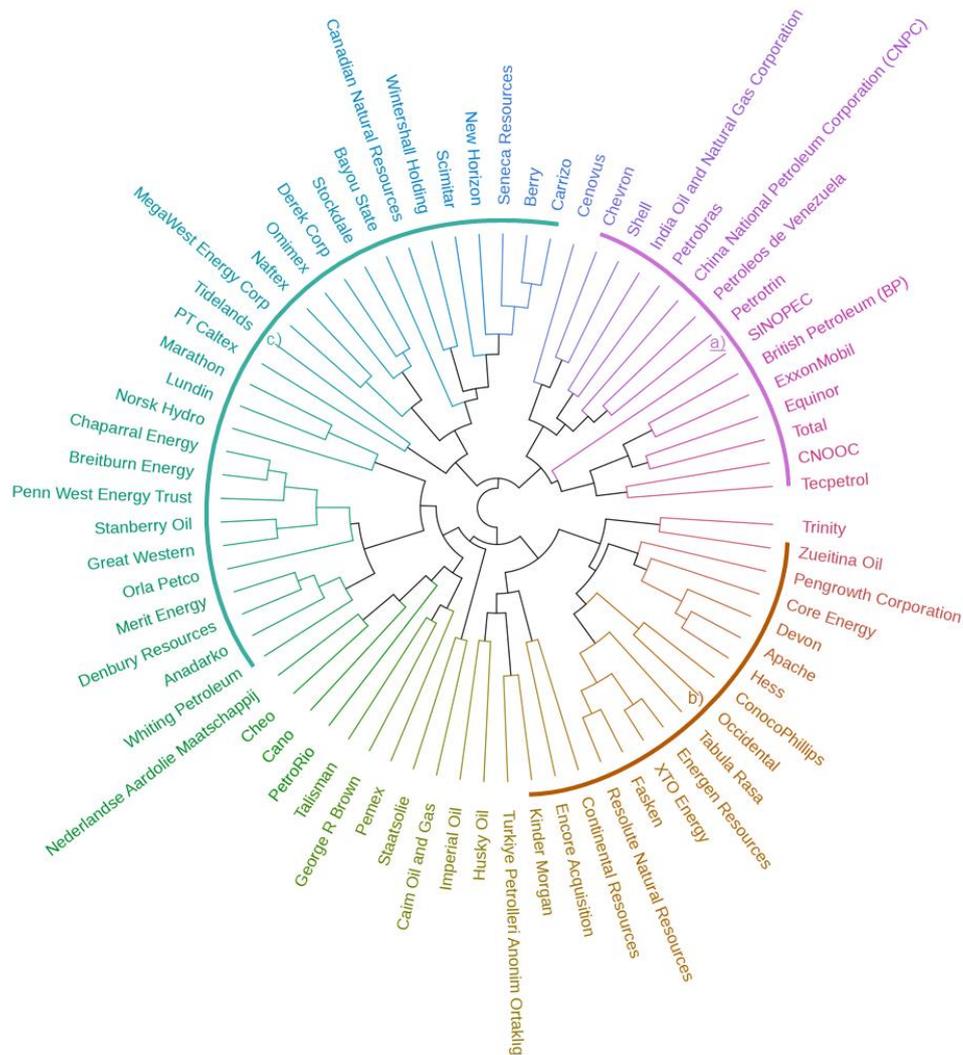


Figura 63. Clusterização das empresas identificadas com projetos de EOR e representação na forma de dendrogramas, com a identificação de três grupos principais de empresas

5.3.2. Obtenção de reservatórios análogos

A busca por campos análogos é uma estratégia vital em estudos de caracterização de reservatórios e planejamento de desenvolvimento, pois permite uma previsão mais precisa de comportamentos e possíveis desafios. A seguir, apresentamos uma análise detalhada dos campos análogos ao campo de Serra, com foco nas características geológicas e operacionais que os tornam comparáveis.

5.3.2.1. Visualização da interrelação das variáveis via PCA

Com o intuito de explorar a interrelação geral entre as variáveis da base de dados de busca e as respectivas importâncias no espaço n-dimensional de variabilidade dos dados, a aplicação do método PCA ao conjunto de dados foi realizada de forma bidimensional (duas *components* principais). Em decorrência da existência de dados faltantes no conjunto de dados, removeram-se as amostras que apresentavam incompletude dos dados. Foram realizadas duas rodadas de PCA, conforme mostra a Figura 64, que ilustra a relação das duas primeiras *components* principais com as variáveis do conjunto de dados. Primeiramente, em a), com um subconjunto das variáveis, em que duas *components* permitiram reter 23,19% da variabilidade do dado. Sequencialmente, em b), com o range total de variáveis, permitindo reter 55,67% da variabilidade do dado. Nota-se relações consolidadas na literatura entre as variáveis. Dentre variáveis que apontam para direções similares, encontram-se profundidade e temperatura, porosidade, permeabilidade e ambiente de sedimentação fluvial; para direções opostas tem-se os pares de variáveis °API e viscosidade, bacia de margem passiva e bacias de ambientes tectônicos compressoriais.

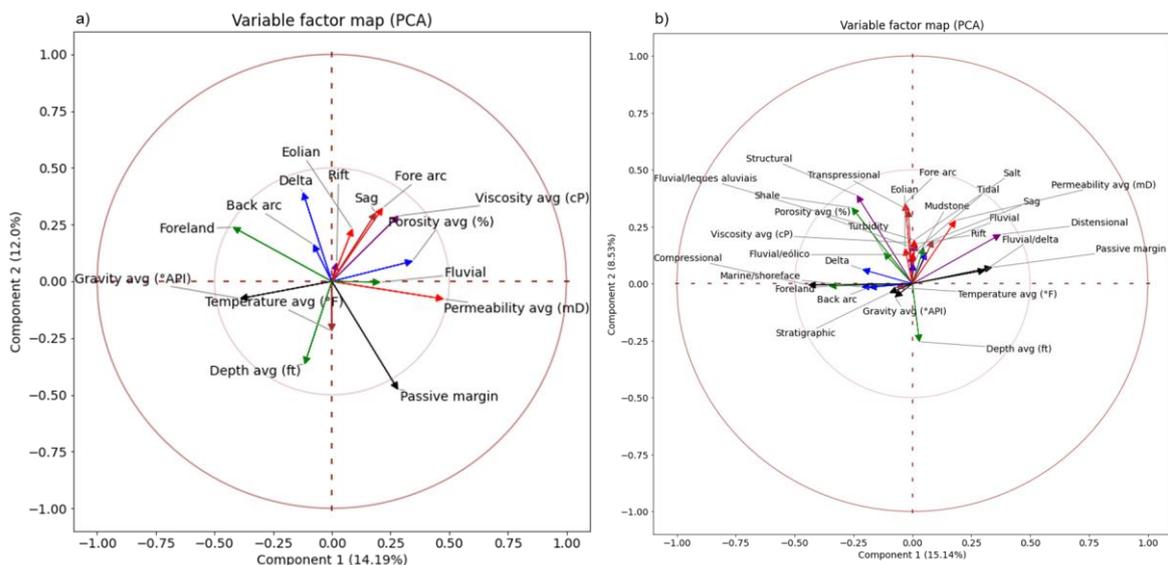


Figura 64. Visualização da relação entre as *components* 1 e 2 com as variáveis do conjunto de dados de campos análogos.

5.3.2.2. Comparação de propriedades numéricas do conjunto de análogos inicial com o reservatório alvo

A aplicação da sequência proposta para obtenção de campos análogos para Campo de Serra da Bacia Potiguar permitiu a obtenção de um subconjunto de campos candidatos a análogos ao reservatório de interesse. Foi realizado corte dos campos que apresentaram medida de similaridade global maior que 0.9 em relação aos reservatórios de Serra, seja o compartimento de óleo pesado ou de óleo leve. A esse conjunto de campos foi designado o nome de candidatos a análogos. Conforme mostra a Figura 65, pode-se comparar as características dos reservatórios desse grupo tanto com as do reservatório de interesse quanto com a base de dados geral utilizada. Observa-se que os dados do grupo de candidatos se encontram com até meio desvio padrão de deslocamento em relação aos reservatórios de interesse.

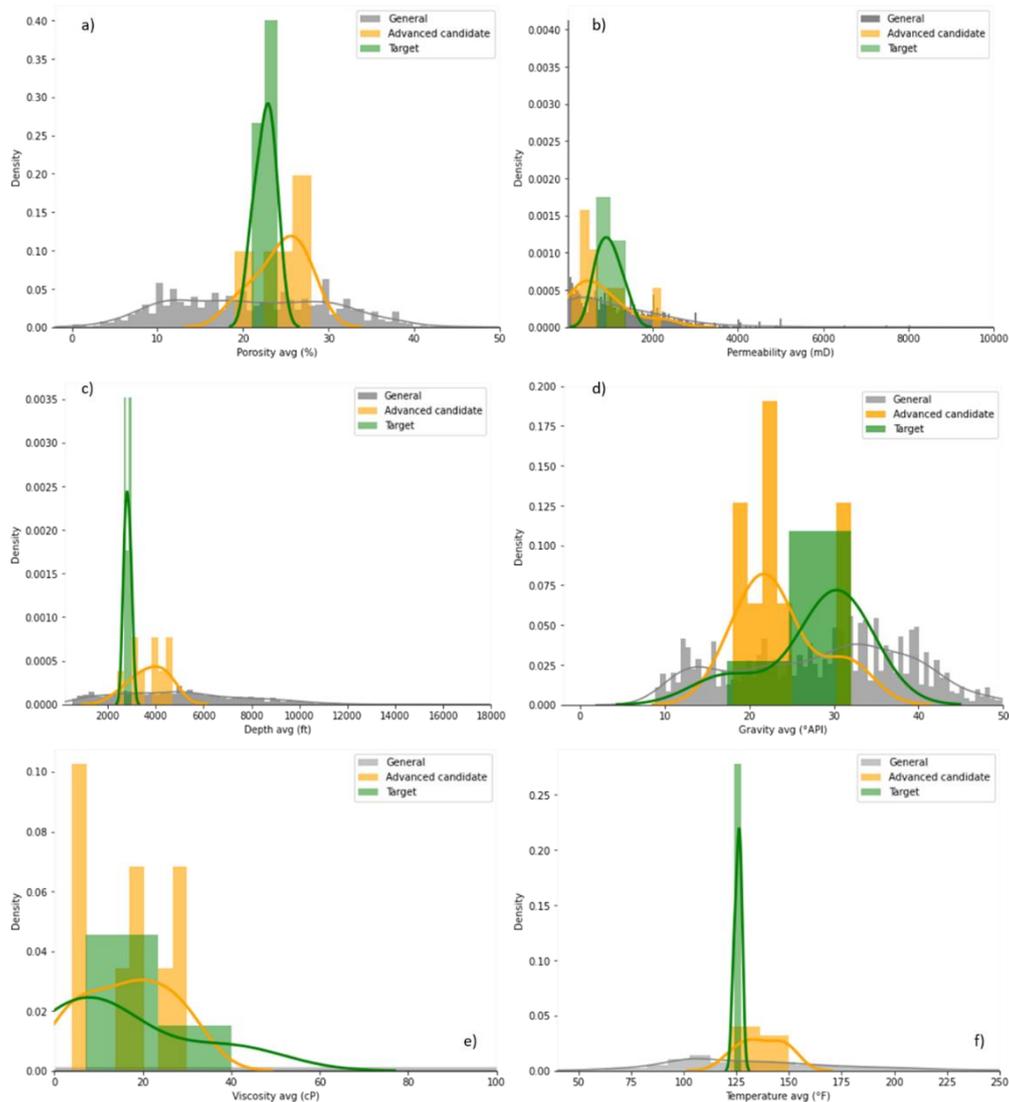


Figura 65. Histogramas comparando os dados do campo de interesse e do conjunto de candidatos a análogos com a base de dados geral para os valores médios das variáveis numéricas a) porosidade, b) permeabilidade, c) profundidade, d) °API, e) viscosidade e f) temperatura de cada reservatório.

Tal comparação também é visível nos scatterplots da Figura 66. Verifica-se que os dados referentes à base mundial de projetos de EOR em geral assumem grande dispersão dos dados, apresentando amplos ranges de variação e representando uma gama variada de reservatórios com diferentes propriedades: a porosidade varia de 4 a 40%, com outliers de até 60%; a permeabilidade média de 0.1 a valores próximos de até 10000 mD; a profundidade dos reservatórios de 500 ft até em sua grande maioria 11 mil ft, com outliers de até 18000 ft; °API de 10 até mais de 50; e viscosidades de 0,1 cP para até mais de 100000 cP para óleos ultra pesados. Verifica-se que os pontos referentes a reservatórios que foram classificados como candidatos possuem range estritamente menor de

dispersão para as variáveis, estando próximo dos ranges referentes às propriedades do campo de interesse.

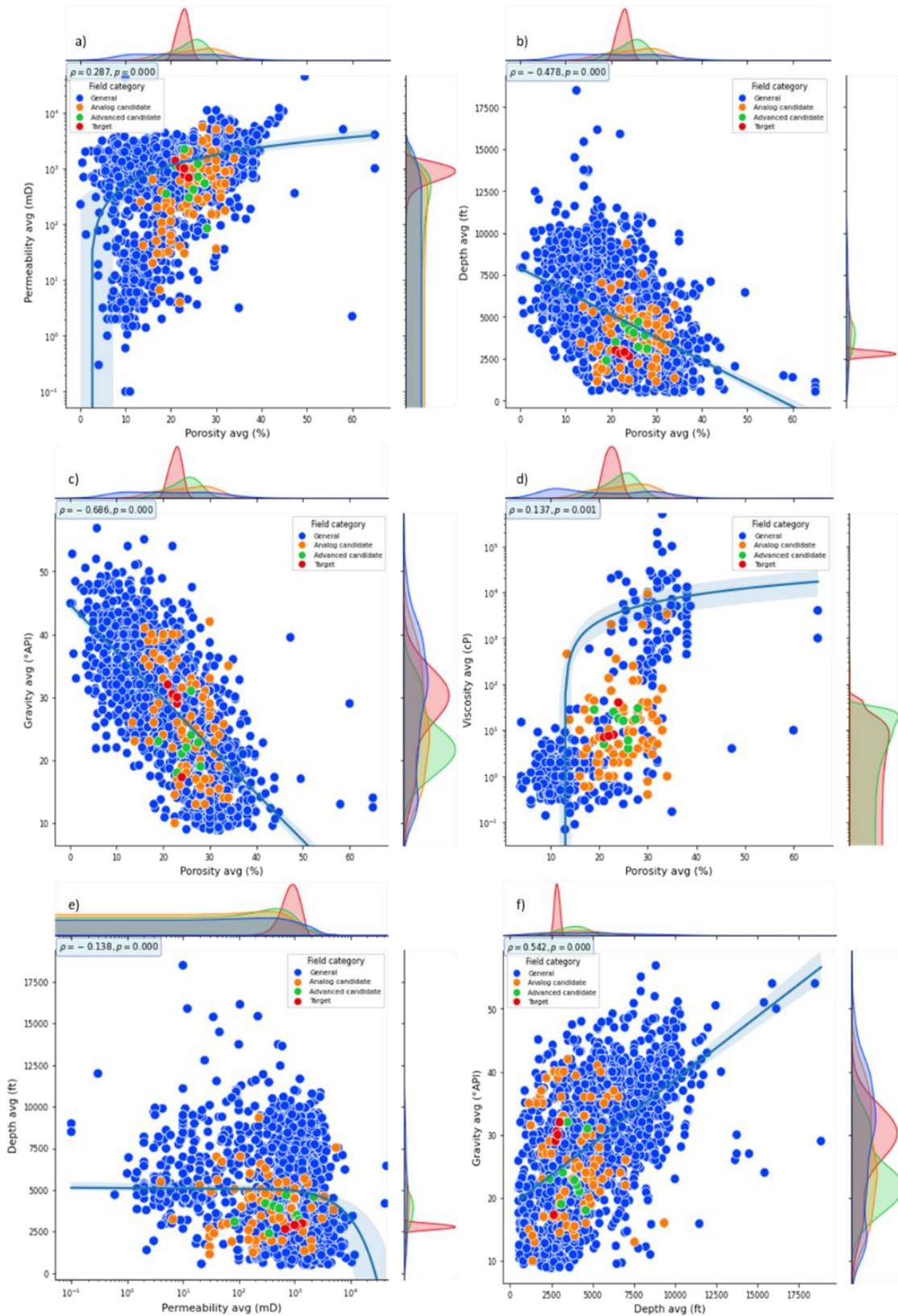


Figura 66. Scatterplots para as variáveis numéricas permeabilidade, porosidade, profundidade, °API e viscosidade do óleo. Os pontos são categorizados quanto a se o dado é referente ao

reservatório de interesse (em vermelho), se faz parte do conjunto de candidatos avançados (em verde), se faz parte do conjunto inicial de candidatos (em laranja), ou se da base de dados inicial.

A Figura 67 sintetiza a distribuição dos valores das propriedades para os campos da base de dados geral, de candidatos análogos e do reservatório de interesse. Verifica-se sobreposição de parte dos polígonos referentes a candidatos análogos com os do reservatório de interesse. Nota-se que para °API há uma menor concentração das curvas referentes aos análogos, devido ao fato de se realizar a busca de análogos para óleo pesado e leve. Já para a variável viscosidade tal dispersão não é visível na visualização por conta de os valores de viscosidade dos análogos e do reservatório de interesse serem da ordem de 10 a 50 cP, números pequenos frente à escala logarítmica com viscosidades de até 100000 cP. Para temperatura, conforme também observado no histograma da Figura 8, nota-se que os candidatos a análogos apresentam valores em geral maiores de profundidade, e em especial, temperatura que o reservatório de interesse. Tal fato é decorrente de reservatórios mais profundos tenderem a apresentar temperaturas maiores, podendo, contudo ser influenciado por aspectos relativos a gradientes geotérmicos.

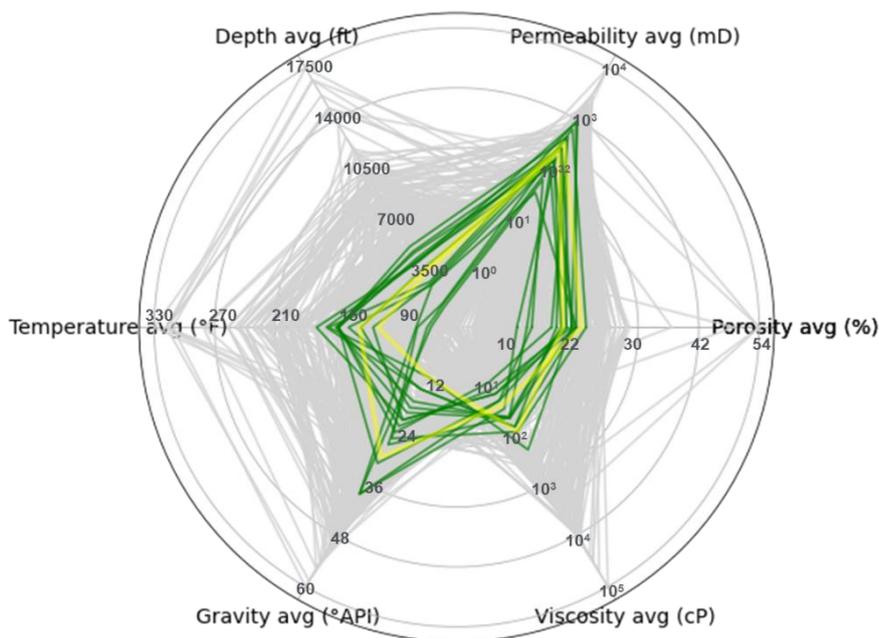


Figura 67. Distribuição dos dados, no formato de polígonos, de acordo com as variáveis numéricas utilizadas. Os dados são categorizados quanto a se é do reservatório target (em amarelo), candidato a análogo (em verde) ou da base de dados geral (em cinza). As variáveis permeabilidade e viscosidade possuem escalas logarítmicas.

5.3.2.3. *Clusterização hierárquica dos reservatórios candidatos a análogos*

A Figura 68 traz a clusterização hierárquica do conjunto de campos levantados como candidatos a análogos ao reservatório de interesse. Todos os reservatórios presentes nesse grupo são areníticos, com permeabilidades altas a médias. É destacada na Figura 68 a formação de diferentes agrupamentos de reservatórios:

1. Região a) pontilhada em azul, com agrupamento de campos que se localizam predominantemente em ambiente tectônico compressional, como bacias sedimentares do tipo foreland, back arc (retroarco) e forearc (antearco). Apesar do estilo estrutural nitidamente oposto, parte desses campos possuem características semelhantes ao de Serra como características permoporosas e propriedades do óleo e que possuem maior influência na dinâmica de reservatórios.
2. Região b) com pontilhado em laranja de grupos de reservatórios que possuem ambiente tectônico predominantemente distensivo, com bacias sedimentares de margem passiva, intracratônicas ou rift. A maior parte dos reservatórios com ambiente de sedimentação fluvial caracterizam-se por estarem associado a esse grupo. Campos desse grupo passaram a serem classificados como “Candidatos avançados a análogos” dada a maior a similaridade com o reservatório alvo de Serra.

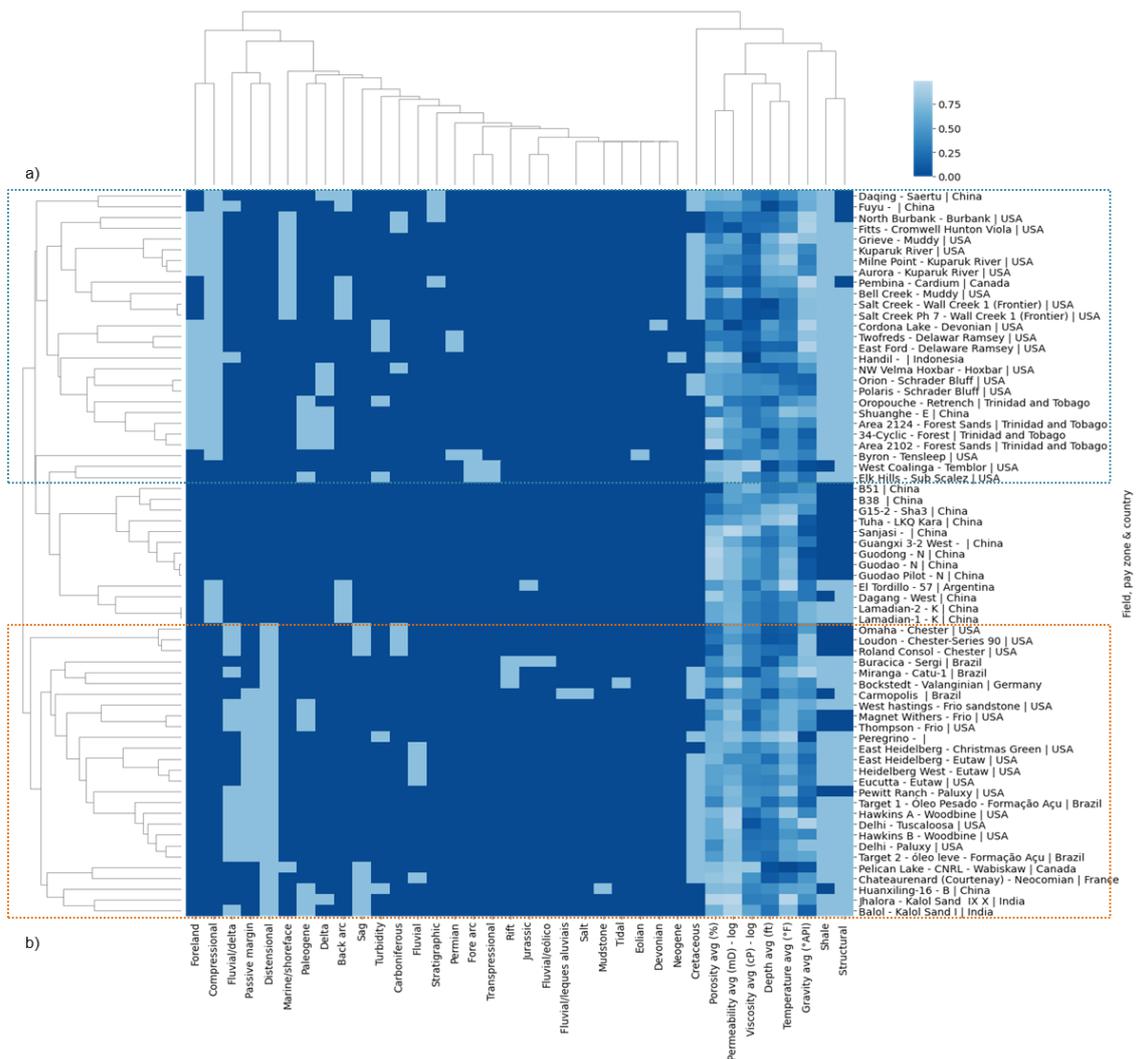


Figura 68. Matriz de clusterização hierárquica obtida utilizando agrupamento aglomerativo para o conjunto de reservatórios candidatos a análogos ao campo de Serra. Cada linha se refere a um campo da base de dados de análogos candidatos, e as colunas a cada variável considerada na busca de análogos.

Conforme a Figura 69, nota-se que em relação apenas a variáveis numéricas o campo de Serra apresenta maior similaridade a reservatórios com diferentes contextos tectônicos e ambientes de sedimentação. Dessa forma, a inclusão de variáveis categóricas e relativas a ambiente de sedimentação e ambiente tectônico possibilita um ganho em permitir considerar aspectos mais abrangentes na definição de campos análogos.

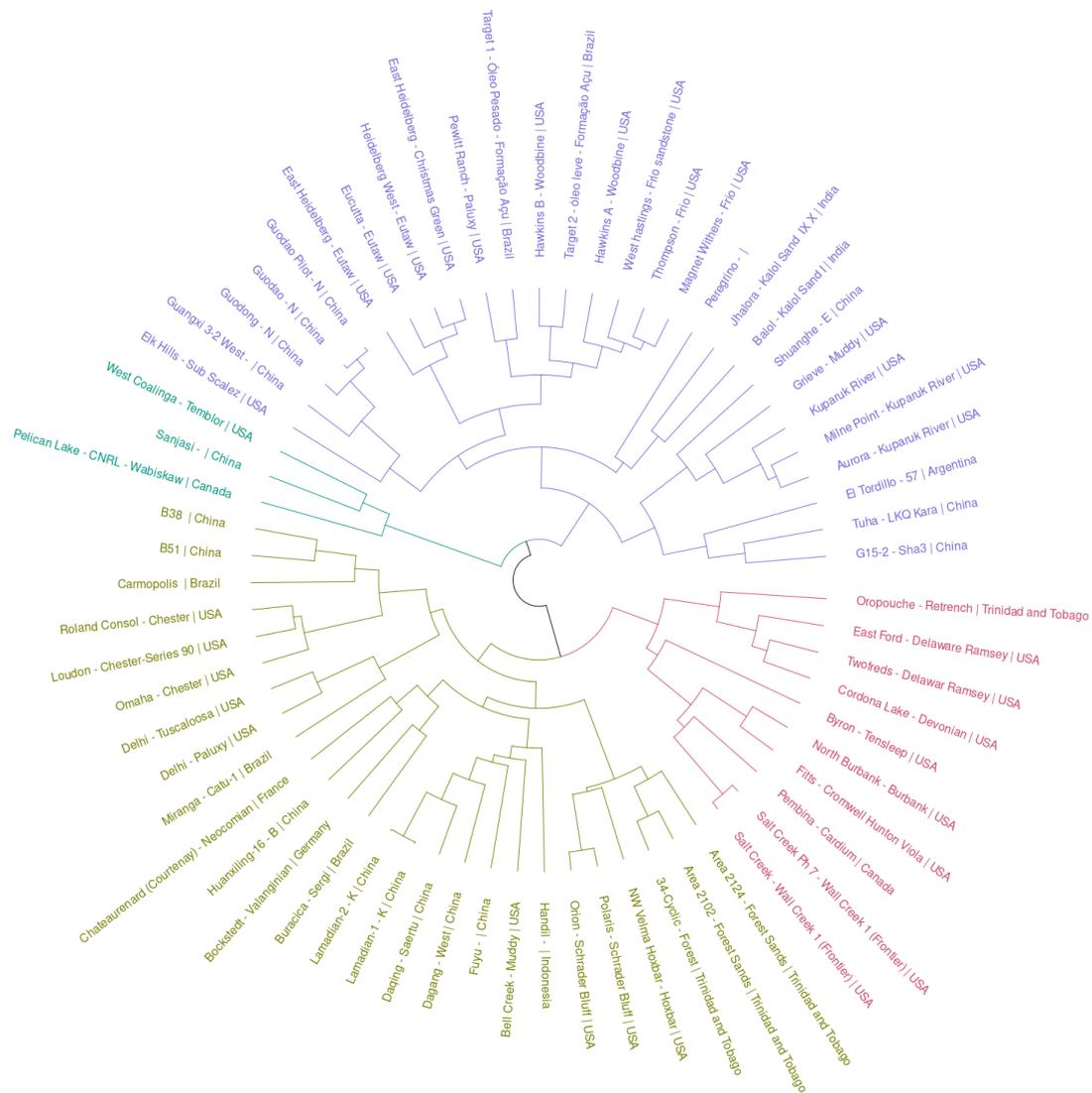


Figura 69. Dendrograma circular obtido a partir da clusterização hierárquica em relação às variáveis numéricas dos campos candidatos a análogos.

5.3.2.4. Comparação qualitativa do conjunto de análogos avançados obtido

Os campos analisados como mais similares em relação ao objetivo definido se localizam em cinco principais bacias sedimentares no mundo, localizadas nos EUA, Índia e França. A Figura 71 traz as famílias de reservatórios análogos internacionais semelhantes ao campo de Serra, considerando tanto as variáveis numéricas quanto categóricas definidas como prioritárias.

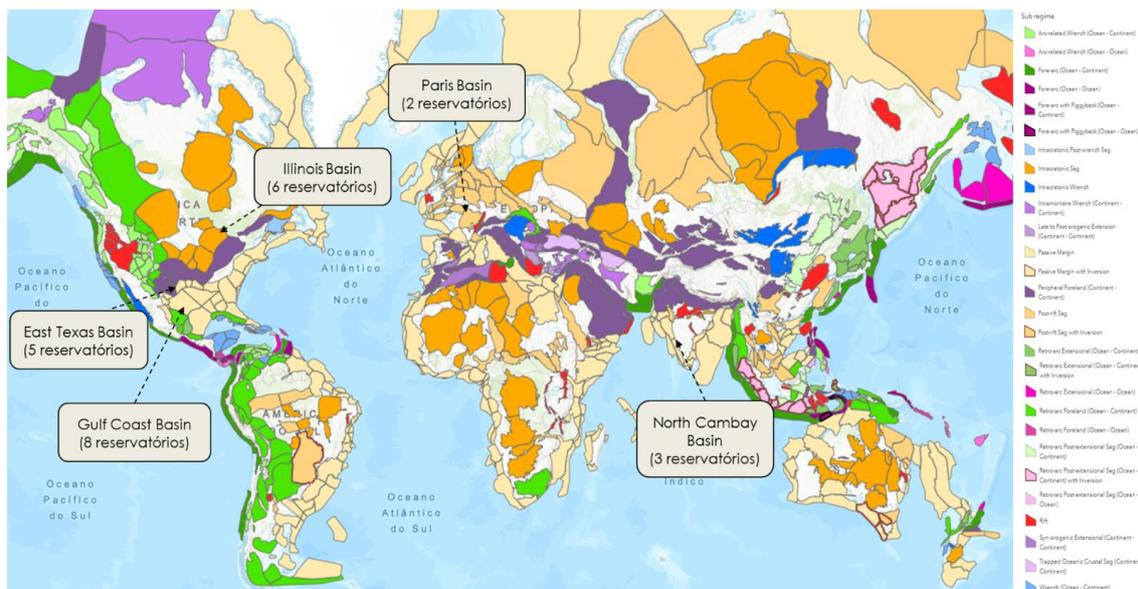


Figura 70. Grupos de reservatórios análogos internacionais semelhantes ao campo de Serra. Mapa extraído e adaptado de ArcGis (2022).

Das famílias de campos obtidas, buscou-se analisar geologicamente e de maneira qualitativa três campos tomados como exemplo, localizados nos EUA nas bacias da Costa do Golfo, do East Texas, e do Illinois, conforme sintetizam os quadros comparativos das Figuras 71, 72 e 73. Nota-se correspondência das características geológicas nos principais níveis para o campo de Serra, tendo como contexto o impacto no desempenho e desenvolvimento de estratégias de produção. Os três campos tomados como exemplos e possuem ambiente tectônico de bacias distensivas, ambiente de sedimentação predominantemente flúvio deltaico, e constituição arenítica da formação com sobrecrecimento de feldspato e sobrecrecimento de carbonatos. Há também diferenças importantes em relação a (i) aspectos específicos como qualidade do reservatório dos campos de Hawkins e Loudon, que são predominantemente quartzo-arenitos e (ii) influência halocinética para o campo de Hawkins, em que há ocorrência de diápiros de sal e significativa presença da região do reservatório em decorrência disso.

Conforme a Figura 72, o campo de West Hastings apresenta características geológicas e tectônicas de bacias muito semelhantes ao campo de Serra. Ambos estão localizados em bacias de margem passiva e em ambientes pós-rifte. A tectônica distensiva é proeminente nos dois campos, com presença de falhas lítricas e rollovers. No que tange ao ambiente deposicional, ambos são fluviodeltaicos, embora o campo de West Hastings North tenha uma influência de ondas

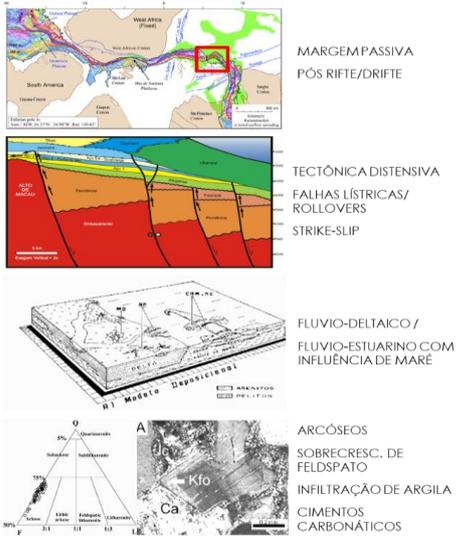
mais pronunciada. Relativamente à qualidade do reservatório, ambos os campos possuem rochas arenosas com sobrecrecimento de feldspato. O campo de West Hastings destaca-se por seus litarenitos feldspáticos e arcóseos líticos, e pela presença de cimentos carbonáticos.



Figura 71. Quadro comparativo de características geológicas do campo de Serra com o campo de West Hastings, da bacia da Costa do Golfo.

O campo de Hawkings, como o campo de Serra, situa-se numa bacia de margem passiva em um ambiente pós-rifte. Conforme a Figura 73, ambos os campos apresentam tectônica distensiva. No entanto, Hawkings difere pela sua estruturação complexa halocinética. Quanto ao ambiente deposicional, enquanto Serra exibe características fluvio-estuarinas com influência de maré, Hawkings tem um domínio fluvial mais acentuado. No que concerne à qualidade do reservatório, Hawkings é notável pelos seus quartzarenitos e sobrecrecimento de quartzo.

CAMPO DE SERRA – BACIA POTIGUAR



CAMPO DE HAWKINGS – EAST TEXAS BASIN

AMBIENTE TECTÔNICO
CONTEXTO ESTRUTURAL
AMBIENTE DEPOSICIONAL
QUALIDADE DE RESERVATÓRIO

MARGEM PASSIVA
 PÓS RIFTE/DRIFTE
 HALOCINÉSE

TECTÔNICA DISTENSIVA
 ESTRUTURAÇÃO COMPLEXA
 HALOCINÉTICA

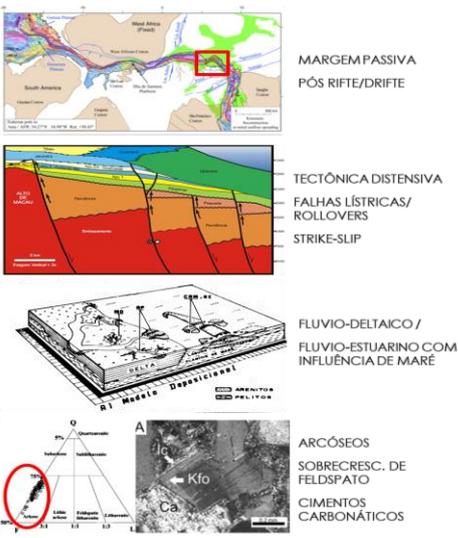
FLUVIO-DELTAICO
 DOMÍNIO FLUVIAL

QUARTZARENITOS
 SOBRECRESC. DE QUARTZO
 INFILTRAÇÃO DE ARGILA
 CIMENTOS CARBONÁTICOS

Figura 72. Quadro comparativo de características geológicas do campo de Serra com o campo de Hawkins, da bacia do East Texas

O campo de Loudon, embora localizado em uma bacia intracratônica, compartilha semelhanças notáveis com Serra, conforme compara a Figura 74. Ambos possuem tectônica distensiva, e seus ambientes deposicionais são fluvio-deltaicos com influências estuarinas e de maré. Na qualidade do reservatório, o campo de Loudon é caracterizado por quartzo-arenitos com sobrecrecimento de quartzo, uma diferenciação notável quando comparado ao predominante arcóseo em Serra.

CAMPO DE SERRA – BACIA POTIGUAR



CAMPO DE LOUDON – BACIA DO ILLINOIS

AMBIENTE TECTÔNICO
CONTEXTO ESTRUTURAL
AMBIENTE DEPOSICIONAL
QUALIDADE DE RESERVATÓRIO

SAG/
 INTRACRATÔNICA

TECTÔNICA DISTENSIVA
 TRAPAS ESTRATIGRÁFICAS
 PREENCHIMENTO DE VALES INCISOS

FLUVIO-DELTAICO /
 FLUVIO-ESTUARINO COM
 INFLUÊNCIA DE MARÉ

QUARTZO-ARENITOS
 SOBRECRESC. DE QUARTZO
 CIMENTOS CARBONÁTICOS

Figura 73. Quadro comparativo de características geológicas do campo de Serra com o campo de Loudon

de Loudon, da bacia do Illinois.

Com base nas Tabelas 5 e 6, observa-se que todos os três campos tomados como exemplos de análogos possuem longos períodos de produção e desenvolvimento, com mais de 50 anos, o que auxilia em realizações de projeções futuras para o de Serra. Em termos de porosidade e permeabilidade, Hawkins apresenta os valores mais altos, sugerindo um potencial maior para a produção. Já Loudon, por outro lado, apresenta os valores mais baixos, e que dadas as demais características indicam um possível desafio para a produção. Comparando as variáveis categóricas (Tabela 6), os campos são muito semelhantes em mecanismos de produção primária, ambiente de sedimentação, estilo estrutural, tipo de selo, tipo de bacia e tipo de trapa. As variações na qualidade do reservatório e composição primária, no entanto, são considerações podem ser críticas ao planejar determinadas estratégias de produção e métodos de EOR.

Tabela 6. Comparativo do campo de Serra com exemplos de análogos em relação a variáveis numéricas.

Campo/reservatóri	Serra	Hawkins	Hastings	Loudon
o				
Pay zone	Açú	Woodbine	Frio sandstone	Chester
Bacia	Potiguar	East Texas	West Gulf Coast Tertiary	Illinois
Coordenadas		32.5884735, 95.2041135	- 43.3739582, 76.1524252	- 39.1388618,- 88.8357539
Anos de produção/ desenvolvimento	-	Mais de 50	Mais de 50	Mais de 50
Litologia dominante	Arenito consolidado	Arenito consolidado	Arenito consolidado	Arenito consolidado
Porosidade média (%)	25	26	28	21
Permeabilidade média (mD)	900	1850	730	250
Profundidade (m)	980	1202	1476	800
Temperatura (°C)	53	61	65	49
Net pay (ft)	30	33	28	34
Grau API	17 30	13 a 24	31	30
Viscosidade (cP)	40 8	45	11	6

Tabela 7. Quadro comparativo do campo de Serra com exemplos de análogos em relação a variáveis categóricas.

Campo/reservatório	Serra	Hawkins	Hastings	Loudon
Mecanismo de produção primária	Aquífero/descompressão	Aquífero	Aquífero	Aquífero
Ambiente de sedimentação	Fluvial/delta/estuário	Fluvial/delta	Fluvial/delta	Fluvial/delta
Estilo estrutural	Distensional	Distensional	Distensional	Distensional
Tipo de selo	Folhelho	Folhelho	Folhelho	Folhelho
Tipo de bacia	Margem passiva	Margem passiva	Margem passiva	Margem passiva
Tipo de trapa	Estrutural/Mista	Estrutural/Mista	Estrutural/Mista	Estrutural/Mista
Idade geológica	Cretáceo	Cretáceo	Paleógeno	Carbonífero
Qualidade de reservatório	Sobrecre. feldspato Infiltração de argila Cimentos carbonáticos	Sobrecre. feldspato Infiltração de argila Cimentos carbonáticos	Sobrecre feldspato/quartzo Cimentos carbonáticos Dissolução de líticos	Sobrecre. de quartzo Cimentos carbonáticos
Composição primária	Arcóseo	Quartzarenitos	Litarenitos feldspáticos/ arcóseos líticos	Quartzarenitos

6. DISCUSSÕES

Com base no estudo desenvolvido, verifica-se que os diversos métodos de recuperação constituem-se em uma ferramenta estratégica e de grande importância para um pleno desenvolvimento da produção de reservatórios de petróleo. Os resultados além de confirmarem informações já respaldadas pela literatura no que concerne engenharia de reservatórios, também permitiram trazer novas percepções acerca de demais aspectos técnicos, econômicos e também estratégicos de aplicação das empresas, mostrando-se um campo oportuno a aplicação de técnicas de ciência de dados em tal problemática.

6.1. Quanto à metodologia empregada

Os resultados apresentados demonstraram coesão entre padrões observados a partir dos mapas SOM para com os modelos reais e teóricos relativos a EOR. Além disso, as informações obtidas demonstraram que a integração e tratamento das diferentes fontes de dados foi realizada de maneira eficiente, permitindo constituir um dataset amplo e coeso acerca de métodos EOR utilizados em escala mundial. O uso de atributos sintéticos, como tamanho

estimado da garganta de poro e mobilidade do óleo, permitiu distinguir classes de métodos de EOR e trazer uma análise mais aprofundada dos parâmetros técnicos e suas implicações. Além disso, com o uso das técnicas de visualizações de dados foi possível analisar a dinâmica dos métodos de EOR de maneira integrada, permitindo compreender melhor as aplicações, vantagens e limitações técnicas de cada método de EOR, e comparar uns com os outros.

As limitações da metodologia adotada, no entanto, incluíram aspectos como a dificuldade para a busca, aquisição, coleta, integração, limpeza e checagem dos dados e ocorrência de dados faltantes para diversas das variáveis, e que muitas vezes reduz o universo de projetos de EOR dentro de uma determinada análise. Para mitigar tais problemas problema, foram utilizadas diversas fontes que permitissem complementariedade em relação às variáveis e dados presentes e foram implementadas etapas de pré-processamento automatizado dos dados reduzindo o tempo dispendido. Além disso, há limitações na utilização do SOMs tanto por na representação gráfica das component plots para cada neurônio ser atribuído um valor estatisticamente próximo ao dos projetos representados, como por suavizar valores extremos. Como exemplo, não foi possível visualizar de maneira clara nos mapas reservatórios com óleos de °API abaixo de 11 ou acima de 40 e permeabilidades abaixo de 5 mD ou acima de 2 mil mD e foram feitas generalizações com base no histórico e padrões observados. Para tanto, a análise dos dados foi realizada de forma conjunto com outras visualizações e inspeção dos dados e validação na literatura.

6.2. Quanto aos resultados e padrões identificados

Com base nos resultados obtidos, é possível notar que variáveis de diferentes naturezas demonstram impacto na escolha, aplicação e sucesso técnico, econômico e tecnológico dos métodos de EOR. Tais variáveis indicam impactos que variam de acordo não apenas com as propriedades técnicas dos reservatórios e o método de EOR em questão, mas também com aspectos locais, comerciais e temporais. Essa análise permite esclarecer parte da complexa e intrincada rede de fatores que influenciam na aplicação de métodos de EOR em escala global. Tem como principal contribuição inovadora a análise gráfica de variáveis não necessariamente técnicas, mas que também

demonstram ter impacto na escolha, aplicação e desenvolvimento de métodos de EOR.

6.2.1. Influência de variáveis técnicas nos métodos de EOR

6.2.1.1. Propriedades básicas dos reservatórios

A análise de propriedades básicas dos reservatórios indica a influência desses parâmetros na escolha e aplicação de método de EOR. Identificou-se nos resultados padrões de amplo consenso na literatura (DONG et al., 2019; LIU et al., 2020; MOKHEIMER et al., 2019; YUAN; WOOD, 2018; YUAN; WANG, 2018) como a aplicação de métodos térmicos de injeção de vapor em reservatórios com (i) óleos pesados, com o intuito de aquecer a formação in-situ com expansão dos fluidos e aumentar a fluidez do óleo, diminuir a saturação residual, promover molhabilidade a água; (ii) formações areníticas e porosidades e permeabilidades altas, de forma a se ter uma capacidade de fluxo mínima de óleo; (iii) profundidades baixas, de maneira que se tenha menor perda na transmissão do vapor; (iv) pequeno número de projetos *offshore*; (v) altas densidade de poços produtores e injetores, possibilitada também pelas baixas profundidades e com o intuito de se ter maior eficiência de varrido, dentre outros. Diferentes padrões também foram identificados para os subtipos de injeção de vapor, com SAGD sendo aplicado em reservatórios óleos extra-pesados, com formações menos espessas ou carbonáticas e em campos que muitas vezes já apresentaram outros métodos de recuperação (DONG et al., 2019; HASHEMI-KIASARI et al., 2014). Para a aplicação de ISC, confirmaram-se tanto similaridades em relação a processos de injeção de vapor, a exemplo de aplicação em óleos em geral viscosos, como também diferenças quanto, por exemplo, referentes à aplicação em condições onde processos de injeção de vapor em geral não se mostram economicamente viáveis, como um maior range de profundidade de reservatórios, porosidades e permeabilidades menores e variação litológica (JI-HONG et al., 2020; MILLER, 1995; MOKHEIMER et al., 2019; TURTA et al., 2007).

Em relação a métodos de injeção de gás, verificou-se que são aplicados em reservatórios com características gerais de: (i) baixos valores de porosidade e permeabilidade, como forma de evitar segregação gravitacional do gás; (ii) viscosidades baixas a intermediárias; (iii) valores médios a altos de profundidades e temperaturas, com o intuito de fornecer condições mínimas para manutenção de pressão no reservatório sem fraturamento da formação. Tais padrões são corroborados por literatura já estabelecida na área (ALVARADO; MANRIQUE, 2010b; AWAN, 2008; LIU et al., 2020; TABER; MARTIN; SERIGHT, 1997). Várias distinções dentro desse grupo puderam ser notadas tanto quanto ao tipo de gás injetado, como também em relação à miscibilidade e continuidade ou alternância do processo de injeção, também abordadas por trabalhos da área (AWAN, 2008; BELAZREG; MAHMOOD, 2020; BRODIE et al., 2012; GBADAMOSI et al., 2018; JIN, 2017, p. 2; KOKAL; AL-KAABI, 2010; LI et al., 2019).

Para o emprego de métodos químicos os principais padrões em relação às características básicas dos reservatórios foram valores intermediários de porosidades, permeabilidades, °API e viscosidades; valores baixos a médios profundidades e temperaturas dos reservatórios, e arenitos litologia dominante com menor número de aplicações em carbonatos. Diversas características particulares de aplicação dos diferentes produtos químicos também foram observadas e que são validadas pela literatura (BELHAJ et al., 2020, 2020; DELAMAIDE; DEGRE, [s.d.]; GBADAMOSI et al., 2019; SHENG, 2010). Os métodos de EOR classificados como “Outros”, que englobam técnicas em sua maioria recentes, possuem grande dispersão e baixa concentração de seus pontos/dados.

6.2.1.2. *Ambiente offshore*

O método de EOR com maior número de projetos *offshore* catalogados foi a injeção de gás hidrocarboneto (miscível/imiscível), seguida pela injeção de polímero e injeção de CO₂. Os primeiros projetos em escala de campo de aplicação de EOR *offshore* se deram principalmente no Mar do Norte, associados a campos vizinhos com elevada produção de gás disponível para injeção. A injeção *offshore* de gás já ganhou elevada maturidade tecnológica,

tendo diversas projetos ao redor do mundo. Já a aplicação de métodos químicos *offshore* começou mais recentemente, podendo ser uma opção viável geralmente onde não há disponibilidade de fonte de gás para injeção. Como exemplos de projetos de injeção *offshore* de polímero, destacam-se os campos de Dos Cuadras, no Golfo do México, Bohai Bay, na China, Captain, no Mar do Norte e Dalia, na Angola. Apesar de casos *offshore* bem sucedidos, verifica-se que ainda persistem desafios relacionados a fatores como: (i) altas profundidades, temperaturas e salinidades, que podem levar à degradação e perda da efetividade dos polímeros e surfactantes; (ii) grande espaçamento entre os poços, exigindo um maior volume e conseqüente maior custo de produtos químicos injetados; (iii) dificuldades logísticas e de espaço em plataforma para o transporte e solubilização de grandes volumes de soluções; (iv) problemas na separação água/óleo devido à baixa tensão interfacial e alta viscosidade dos fluidos produzidos com quantidades remanescentes dos fluidos injetado, acarretando dificuldades no processo de flotação e limitando a coalescência. Nesse contexto, a injeção de polímero possui um maior avanço em aplicações *offshore*, quando se comparado, a ASP/SP por conta do menor consumo de produtos químicos, maior maturidade e menor complexidade de implementação. Por outro lado, há diversos avanços em pesquisa para o desenvolvimento de novas substâncias e aditivos que mostram a possibilidade de superar algumas dessas barreiras para a utilização de *offshore* de novos métodos químicos (BONDOR, 2005; KANG; LIM; HUH, 2016; RANEY et al., 2012; REKSIDLER et al., 2015).

Se comparada às outras classes de métodos de EOR, a injeção de vapor apresenta quantidade pequena de projetos *offshore* devido a fatores como perda de calor durante a injeção, menor densidade de poços para se obter um varrido eficiente, desafios relacionados a equipamentos requeridos para aquecimento, pressurização e injeção de vapor e que são difíceis de se manter *offshore*, dentre outros.

6.2.1.3. *Varáveis sintéticas*

Verificou-se ganho de informação na identificação de correlações e padrões ao se criar variáveis sintéticas a partir da combinação de variáveis primárias.

Para a variável sintética garganta estimada de poro verificou-se correlação significativa estatisticamente com as variáveis sintéticas diferença entre saturação inicial e final e com a eficiência estimada de deslocamento; tais correlações estatísticas foram maiores do que as correlações existentes das variáveis porosidade e permeabilidade com os valores de saturação e eficiência estimada de deslocamento. Além disso, com a inclusão da variável da garganta estimada de poro a separação dos dados dos grupos de tipos de gás injetado se verificou mais nítida do que apenas com as variáveis porosidade e permeabilidade. Tais fatos indicam impacto do tamanho da garganta de poro e da mobilidade do óleo na eficiência de deslocamento e recuperação do óleo em processos de injeção de gás, com correlação maior do que variáveis primárias como porosidade e permeabilidade.

Para a variável mobilidade estimada do óleo permitiu-se associar em um mesmo parâmetro propriedades relativas à capacidade de fluxo permitida pela rocha (permeabilidade) com a quantidade inercial de deslocamento do fluxo (viscosidade do petróleo). Com a adição de tal variável observou-se ganho de informações a respeito da aplicação das classes de EOR, obtendo-se, por exemplo, distinção de métodos de EOR nos *boxplots* e separação de grupos de pontos nos *scatterplots*.

6.2.2. Influência de aspectos locais na escolha de métodos de EOR

A partir dos resultados foi possível observar influência de aspectos locais na escolha, aplicação e desenvolvimento econômico de técnicas de EOR, mesmo quando se apresentavam critérios técnicos satisfatórios para a aplicação de EOR. Ainda que a adoção de estratégias de recuperação avançada tenham se difundido mundialmente, nota-se a concentração da aplicação de determinados métodos de EOR em um número reduzido de localidades. Tais concentrações se mostraram ser associadas a fatores desde disponibilidade e custos de fornecimento e transporte de fluido injetante, histórico de produção dos campos, expertise no desenvolvimento de tecnologias por parte das empresas, até o grau de desenvolvimento da indústria e infraestrutura, aspectos regulatórios e políticas de incentivo à produção de recursos naturais remanescentes de cada país.

A injeção de gás hidrocarboneto é um dos métodos de EOR que possui aplicação em um maior número de países e regiões do mundo. Estudos mostram a injeção de hidrocarboneto como apresentando em geral menor eficiência de recuperação se comparada a outros métodos de injeção de gás, como a injeção de CO₂ (GBADAMOSI et al., 2018; LATIL et al., 1980). Contudo, nota-se que muitas empresas decidem por aplicar injeção de hidrocarboneto, em detrimento de outros métodos de EOR que podem ser mais eficientes. Dentre os motivos para isso destacam-se fatores não exclusivamente relacionados a propriedades dos reservatórios, como em geral disponibilidade de gás para injeção obtido de campos vizinhos, maior conhecimento do comportamento e dinâmica do gás no reservatório, menores custos associados a corrosão, dentre outros.

Métodos como injeção de CO₂ e o uso de produtos químicos (polímero e surfactante) apresentam sua aplicação concentrada em menor número de países, como EUA e China, estando associados a fatores como uma rede de gasodutos de transporte de CO₂ e uma indústria química desenvolvida com fornecimento de produtos de química fina a baixo custo, respectivamente. Os métodos térmicos, por sua vez, se mostram em geral associados a localidades que possuam, dentre outros fatores, reservatórios com características específicas, como óleo pesado e elevada capacidade de fluxo poroso.

6.2.2.1. *Aspectos regulatórios – políticas de incentivo*

EUA e Canadá

Nota-se que a aplicação de EOR se deu em diferentes momentos em diferentes países, sendo muitas vezes motivada por políticas energéticas e incentivos fiscais para a aplicação de EOR.. Com os choques do petróleo em 1973 e 1979, países com já grande parte dos campos de petróleo parcialmente depletados, como EUA e Canadá, enxergaram na aplicação dos métodos de EOR a possibilidade de se suprir parte de demanda e segurança energética e maior aproveitamento das reservas. Os EUA são um bom exemplo de como os incentivos políticos afetam o crescimento dos projetos de EOR. Na década de

1980, diante da perspectiva de declínio da produção doméstica de petróleo, o Crude Oil Windfall Profit Tax 1980 deu o pontapé inicial na indústria de EOR dos EUA, reduzindo significativamente sua carga tributária para projetos de EOR de até \$15 milhões de dólares na época (BELAL; CLARK, 1981; CRS, 2006). Em 1990 o congresso americano promulgou a Seção 43 do Internal Revenue Code para fornecer um crédito de investimento, de até 15% e dependente do preço do barril de petróleo, para certos custos que um operador paga ou incorre em projetos de EOR (BECKNER; MARWICK, 1994). Além disso, o Departamento de Energia dos EUA forneceu suporte técnico aos operadores, incluindo ferramentas que auxiliam no *screening* de métodos de EOR. Mais recentemente, o crédito fiscal da seção 45Q dos EUA foi alterado para fornecer uma redução de impostos de US\$ 35/tCO₂ por 12 anos para CO₂ armazenado em operações de EOR (ANDERSON et al., 2021).

Em virtude da regulação com menor centralização nos EUA para uso dos recursos do subsolo, também se destacam iniciativas locais, a exemplo do estado do Texas que fornece uma taxa especial para novos projetos EOR aprovados ou para a expansão dos existentes. A taxa é de 2,3% do valor de mercado da produção (metade da taxa padrão) por 10 anos após a certificação da resposta da produção pela Comissão (RRC, 2022). Em 2007, o Texas também aprovou uma “Advanced Clean Energy—EOR Tax Reduction” para projetos de armazenamento de CO₂-EOR usando fontes antropogênicas (MILLER, 2014). Para promover a implementação de projetos de EOR, em 2003 Wyoming reativou uma redução na taxa de indenizações de 6 para 4 por cento nos primeiros cinco anos de produção incremental proveniente de operações de EOR (BLM, 2018).

No Canadá, Saskatchewan Petroleum Innovation Incentive (SPII) oferece créditos de imposto de produção de royalties/propriedades de petróleo e gás transferíveis para projetos de comercialização de inovação qualificados a uma taxa de 25% dos custos elegíveis do projeto. O principal objetivo desta iniciativa foi permitir aos candidatos a capacidade de testar suas inovações em um ambiente piloto e implantá-las em escalas comerciais maiores, demonstrando viabilidade técnica e econômica e aumentando a conscientização sobre projetos de CO₂-EOR (SPII, 2022). Políticas como essas levaram a um

grande aumento no número de projetos de EOR iniciados nas décadas de 80 e 90, e em especial projetos de injeção de gás e CO₂ a partir dos anos 2000, conforme pode ser observado no gráfico da Figura 11.

Mar do Norte

Em países do Mar do Norte, como Noruega, Reino Unido e Escócia, esforços conjuntos da academia, institutos de pesquisa, governos e empresas levaram a uma estratégia de buscar melhorar a recuperação dos campos por meio da injeção de gás, geralmente proveniente de campos vizinhos. Dado os longos prazos de retorno e a necessidade de envolver os fornecedores de infraestrutura de geração, captura e transporte de CO₂, um anúncio antecipado e proativo de um incentivo fiscal específico para CO₂-EOR pelo governo do Reino Unido enviaria um sinal positivo para a indústria petrolífera e a indústria do CCS. Qualquer incentivo, no entanto, precisaria ser revisado regularmente em função do mercado (ou seja, preços do petróleo) e das condições regulatórias. Na Noruega, a produção de petróleo bruto caiu pela metade de um pico de 181 milhões de metros cúbicos em 2000 para 81 milhões de metros cúbicos em 2019 (NPD, 2021). Diante dessa queda de produção de óleo e do volume de reservas, a agência reguladora da Noruega, NPD, tem incentivado o emprego de EOR e sinalizado a possibilidade de obrigar empresas operadoras a utilizarem métodos de EOR, especialmente nos casos de haver lucro significativo em jogo (HART ENERGY, 2017; NPD, 2020; REUTERS, 2017).

China

Na China existem programas governamentais de redução de taxas de 40% para a produção de óleos pesados, 30% para operações com EOR, 20% para campos de gás com pouco óleo e 30% para operações de óleo e gás em águas profundas (CNPC, 2014; LING, 2014; WEI et al., 2015). O maior projeto de EOR químico foi implantado na Daqing, na China, mesmo em períodos com baixos preços do petróleo, por conta de definições de Estado em um cenário de grande importação de petróleo (LE et al., 2015; SUN et al., 2018). A China apresenta ainda incentivos locais, como na província de Xinjiang, de 30% do

imposto sobre recursos ou 1,5% da receita total do petróleo, para o desenvolvimento de EOR.

Trindade e Tobago

Em 2011 e 2012, o governo de Trindade e Tobago introduziu reduções de impostos para empresas que produzam usando recuperação aprimorada de petróleo de campos maduros, baixando o SPT progressivamente dependendo do preço do petróleo. A Tabela 2 resume as taxas de imposto. A partir de 1º de janeiro de 2022, o governo de Trindade e Tobago irá conceder um crédito fiscal de 30% até um máximo de US\$ 500 mil para empresas que realizarem investimentos em projetos de CCS e EOR (MEEI TRINDAD AND TOBAGO, 2021). Iniciativas como essa são movimentos positivo do Governo no esforço de mitigar as mudanças climáticas e, simultaneamente, aumentar a produção de petróleo com a injeção de CO₂ em projetos de EOR. No entanto, a infraestrutura necessária para capturar, transportar e armazenar o CO₂ pode exigir um desembolso de capital significativo e o custo do CCS foi identificado como o principal desafio para a adoção generalizada dessa tecnologia no país (PWC, 2022).

Índia

Mais recentemente a Índia implementou política de EOR compulsória para os ativos do país. Na última década a Índia apresentou poucas novas descobertas expressivas de petróleo e um consumo crescente de óleo, que possui um papel crucial na economia nacional, o que levou o país a aumentar sua dependência da importação de petróleo (BERA; VIJ; SHAH, 2021; DUTTA, 2020; SAKTHIPRIYA; DOBLE; SANGWAI, 2015). Diante desse cenário de insegurança energética, se tornou uma necessidade o governo indiano constituir políticas que estimulem a implementação de Enhanced Recovery (ER). Para tanto, foram implementados incentivos fiscais de 50 a 75 % para a produção incremental obtida – além do previsto por *benchmark* (BERA; VIJ; SHAH, 2021; DUTTA, 2020; ENERGYWORLD, 2018). Além disso, também estabeleceu-se diretrizes para realização mandatória de *screening* de métodos de EOR para

todos os ativos com produção comercial superior a três anos, com o intuito de verificar o potencial da aplicação de técnicas de EOR e a execução de projetos pilotos e posteriormente em escala de campo (BERA; VIJ; SHAH, 2021; DUTTA, 2020). Existe um enorme potencial de sucesso para a implementação de EOR nos campos da Índia e junto à atual Política de Enhanced Recovery, conforme exigido pelo Governo da Índia, é amplamente esperado que as implementações de triagem de ER ganhem ritmo na região nordeste da Índia.

Indonesia

A Indonesia, hoje um dos países de fontes mais diversificadas em termos de fornecimento de energia, era anteriormente conhecido como o primeiro país a criar o sistema fiscal de Contrato de Partilha de Produção (PSC) na década de 1960 (Lubiantara, 2012) que foi proposto pela Pertamina para lidar com investimentos estrangeiros em negócios de petróleo e gás que estavam começando a florescer durante esse período específico. No entanto atualmente está com número cada vez menor de novas descobertas de campos de petróleo.

A implementação de EOR na Indonesia está limitada ao campo de óleo pesado de Duri, com injeção de vapor, operado pela Chevron. A grande quantidade de reservas no campo encontradas na Bacia Central de Sumatra é um dos principais fatores para tornar o projeto economicamente lucrativo, mesmo com baixo preço do barril e elevado *watercut* envolvido. Em ativos menores empresas preferem utilizar elevação artificial, fraturamento hidráulico, acidificação, otimização da produção e outras técnicas que promovem ganhos mais a curto prazo, e que possui menores custos e riscos envolvidos quando se comparado à implementação de EOR (CHANDRA; RACHMAT, 2018). Diante disso, em meados de 2017, o governo introduziu um novo esquema de prêmio dividido que aborda as complicações na operação de campo, desde complexidades geológicas e de hidrocarbonetos até da estratégia de recuperação utilizada com incentivos diretamente relacionados a: (i) status de produção, (ii) localização (*onshore* vs *offshore*), (iii) profundidade do reservatório, (iv) disponibilidade de infraestrutura (desenvolvidas vs novas fronteiras), (v) tipo do reservatório (convencional vs não convencional), (vi) conteúdo de CO₂ e H₂S, (vii) °API, (viii) conteúdo local, (ix) estágio da produção (recuperação primária,

secundária ou terciária com uso de EOR), (x) função do preço do barril e preço do gás e (xi) relação com a produção acumulada no campo (MINISTER OF ENERGY AND MINERAL RESOURCES; REPUBLIC OF INDONESIA, 2017; ROACH; DUNSTAN, 2018). Isso apresenta um sistema que o governo considera adequado para alavancar o estado dos negócios de petróleo e gás da Indonésia; além disso, a atração de operadores experientes em EOR e outros métodos modernos para investir e desenvolver ativos maduros pode levar o setor a um outro patamar (MCKINSEY, 2020).

Outras iniciativas

Além dessas iniciativas, outros países, como o Brasil, vêm implementando incentivos fiscais para o desenvolvimento de campos maduros. Há a possibilidade da expansão dessas políticas para a implementação de métodos de EOR. Tais políticas a recuperação de recursos remanescentes se destacam por apresentarem diversos impactos positivos para a economia e sociedade nos países em que são implementadas, a exemplo de: (i) estenderem a vida útil dos campos de produção; (ii) aumentar o fator de recuperação; (iii) aumentar a produção de óleo e segurança energética; (iv) viabilização de novos investimentos; (v) garantir a arrecadação de royalties por mais tempo, tanto para a União quanto para estados e municípios; (vi) menor alíquota de royalties, que pode significar a diferença entre a continuidade e o fechamento de suas atividades; (vii) impacto direto na cadeia de fornecedores instalada e desenvolvimento local de novas tecnologias, empreendimentos e novos empregos diretos e indiretos. Por outro lado, verifica-se que em países com grandes reservas e baixos custos de extração e/ou com maiores instabilidades políticas não se vê incentivo para empresas atuarem em campos deletados, que possuem produções menores e diluídas no tempo, com menor payback e TIR.

6.2.3. Influência de variáveis comerciais nos métodos de EOR

6.2.3.1. Preço do barril de petróleo

Os resultados obtidos também sugerem uma influência do preço do barril de petróleo nas tomadas de decisões relativas à implementação de projetos de EOR. Possíveis influências mostram afetar não apenas o início dos projetos e a pesquisa e aplicação de novos métodos de EOR, como também suas continuidades no longo prazo.

A análise que compara os números de publicações científicas no decorrer do tempo com os preços nominais e ajustados do barril de petróleo de cada ano sugere influência do cenário econômico mundial na pesquisa, desenvolvimento e inovação tanto de novos métodos de EOR como também no ganho de maturidade de métodos de EOR já existentes. Ainda que haja influência de diversos outros fatores implícitos, verifica-se também tal associação com o preço do barril ao número de projetos de EOR existentes em determinados períodos, variando conforme o método de EOR.

O preço do barril aparenta apresentar menor impacto no estímulo da aplicação e continuidade de projetos métodos de EOR mais maduros tecnologicamente, como processos de injeção de gás e de vapor; nesses casos outros fatores parecem exercerem um peso maior, como disponibilidade de injetantes, tipo de companhia operadora, políticas regulatórias locais, dentre outros. Verificou-se para injeção de hidrocarboneto baixos valores do preço do barril ajustado no momento do início dos projetos, quando comparado a outros métodos. Tal fato pode ser associado ao motivo de a preços mais altos do barril haver uma priorização para se distribuir e comercializar o gás, monetizando-o, em detrimento de injetá-lo nos reservatórios vizinhos. Para a injeção de CO₂ nota-se relativa independência quanto ao preço ajustado do barril. Conforme afirmam (BABADAGLI, 2020; MANRIQUE; ENGINEERING; MUCI, 2007) essa menor sensibilidade ao preço do barril para a implementação e continuidade de projetos de injeção de CO₂ pode ser associado, em parte, ao fato de esses projetos possuírem fontes próximas já estabelecidas de fornecimento e transporte de gás para a injeção.

Já para métodos de EOR com menor maturidade tecnológica ou mais intensivos em termos de custo, como métodos químicos e novas tecnologias, o preço do barril demonstra maior influência tanto sobre o número de projetos iniciados como também sobre a escala de aplicação (se foram conduzidos projetos pilotos, em escala de reservatório ou em escala comercial de campo).

Diversos autores também sugerem influência do baixo preço do barril principalmente na década de 1980 como um dos motivos para redução da implementação de métodos químicos em escala de campo (BABADAGLI, 2020; HAMA et al., 2014; PUSCH, [s.d.]; SALEH; WEI; BAI, 2014; YUAN; WOOD, 2018). Em períodos de preços baixos do petróleo, há maior competição pelas fontes de financiamento disponíveis e apenas os projetos mais lucrativos, de maior maturidade tecnológica e de menor risco tendem a ser sancionados. Dentre os métodos térmicos, projetos com aplicação de SAGD estão associados a preço de petróleo altos. Métodos de EOR que requerem maiores investimentos (como SAGD), apesar de tecnologicamente maduros, se mostram mais sensíveis a flutuações no preço do barril quando comparados a outros métodos térmicos, como injeção contínua e/ou cíclica de vapor e combustão in-situ.

Verificou-se também correlação estatística significativa do preço ajustado do barril com a profundidade e temperatura dos reservatórios para a aplicação de métodos químicos, especialmente injeção de polímero. Para projetos de EOR com aplicação de métodos químicos, observou-se a tendência de maior profundidade dos reservatórios quanto maior o preço do barril no início do projeto. Contudo, tal correlação do preço ajustado do barril vs profundidade dos reservatórios não apresentou significância estatística relevante para projetos de injeção de gás. Tal fato pode indicar o desafio tecnológico e econômico de se aplicar polímero e surfactantes em profundidades e temperaturas maiores (devido a fatores como degradação e estabilidade termodinâmica) e um ganho de maturidade e avanço da tecnologia, e cuja atratividade econômica e aplicação nesses contextos de elevadas profundidades e temperaturas tende a ser mais alta e estimulada em cenários de preços do barril mais altos.

6.2.3.2. *Porte, segmento e posicionamento de empresas operadoras*

Os resultados indicam possíveis relações do porte, segmento de atuação e posicionamento estratégico das companhias operadoras com fatores não somente geológicos das características dos campos, como também aspectos técnicos e de maturidade tecnológica dos métodos de EOR, além do preço do barril de petróleo.

Verifica-se diferenciação dos diferentes portes de empresas e respectivos posicionamentos estratégicos quanto às características dos campos e tipos de tecnologias de EOR empregadas. Pequenas e médias empresas focam em métodos de EOR que já possuam maior desenvolvimento tecnológico e maior maturidade tecnológica e comprovações de sua eficácia. A aplicação de novos métodos fica limitada a NOCs e IOCs, por novas tecnologias em geral apresentarem maiores riscos e requererem maiores estudos e testes experimentais e pilotos para comprovar suas viabilidades técnica e econômicas para o campo de interesse. Em especial, se comparada a IOCs, nota-se uma adoção maior por parte das NOCs de métodos de EOR que possuem recentes ou que possuam maiores incertezas, como por exemplo LSWF e injeção microbiana. A injeção microbiana possui tentativas de aplicações há décadas em projetos pilotos, porém sua aplicação em larga escala tem se mostrado difícil devido a incerteza nos seus resultados e a necessidade de um longo ciclo de testes e experimentos. Já o emprego de LSWF é considerado recente, porém a partir dos anos 2000 já teve aplicações de escala de campo, com crescimento na sua utilização a partir de 2010. Apesar de ser recente a sua aplicação, LSWF destaca-se por já apresentar aplicações de campo também em ambiente *offshore*, como os campos areníticos de Endicott no Alaska e Snorre, Heidrun e Clair Ridge no Mar do Norte (CHEN et al., 2021). Os campos com projetos de LSWF foram em sua maioria conduzidos por IOCs e em menor número por NOCs, geralmente com fonte próxima de captação de água (como água do mar), e com nenhuma aplicação catalogada realizada por operadoras independentes de pequeno e médio porte.

Boa parte dos campos gigantes do mundo atingiram a maturidade e ainda têm uma quantidade considerável de petróleo deixada para trás para ser recuperada. Há um grande movimento, por parte das grandes operadoras, de venda de ativos tanto de campos com grande porção do óleo recuperável já depletado quanto de campos de pequeno porte. Para as grandes operadoras tais ativos passam a apresentar menores taxas internas de retorno (TIR), se tornando menos atrativos com o passar do tempo quando se comparados a novos ativos e descobertas de reservas. Além disso, no contexto de transição energética e de incertezas quanto à demanda e aos investimentos nas próximas décadas no setor, torna-se uma necessidade as grandes empresas buscarem produzir da

maneira mais célere possível os recursos petrolíferos que possuem em mãos; dessa forma as majors tendem a desinvestir de campos que já tenham a apresentar ganhos marginais, ainda que com lucros no longo prazo. Contudo, apesar de apresentarem perspectivas de menores retornos financeiros (TIRs), a compra de ativos depletados com grandes números de poços tende a apresentar menor risco geológico e de infraestrutura, e assim são opções preferidas por empresas menores. Essas opções possuem a possibilidade de fornecer fontes de receitas significativas com fluxo de caixa distribuído no longo prazo para pequenas empresas (BABADAGLI, 2007; TELES et al., 2015). Tal fato pode ser exemplificado ao se ter observado nas visualizações dos resultados que campos operados por empresas médias e pequenas tendem a empregar maior densidade e menor espaçamento entre poços por meio de estratégias como infill drilling, pertinente principalmente em campos maduros já bastante depletados, como forma de recuperar o óleo remanescente e zonas não varridas. Em campos maduros, o conhecimento acerca do comportamento dos respectivos reservatórios tende a possibilitar a aplicação de EOR com menores incertezas, diminuindo os riscos financeiros e ambientais associados (ALVARADO; MANRIQUE, 2010; MANRIQUE, 2009). Já em campos com pequenas reservas a aplicação de EOR via esquema Huff n Puff, com ciclos de injeção, fechamento e produção em números reduzidos de poços, tem se mostrado uma alternativa capaz de reduzir riscos e incertezas para pequenos produtores.

6.2.4. Similaridades globais entre os métodos de EOR

Para a classe Outros observa-se a formação de três grupamentos distintos e separados, por se referirem a métodos com tipos de reservatórios de aplicação e princípios físico e químicos de ação e funcionamento notavelmente diferentes. Observa-se proximidade e fronteiras de baixa dissimilaridade do microbial com métodos térmicos e químicos. Já da injeção de Foam com gás observa-se proximidade com métodos de injeção de gás, porém com fronteiras de alta dissimilaridade, indicando diferenças substanciais na aplicação desses métodos. Para LSWF nota-se proximidade e fronteira de baixa dissimilaridade com métodos químicos e de injeção imiscível de gás; e proximidade com métodos térmicos, porém fronteiras de alta dissimilaridade.

6.3. Quanto à obtenção de campos análogos

A abordagem integrada de técnicas de visualização e análises de dados, envolvendo clusterização hierárquica, PCA, e histogramas, revelou-se de valor inestimável para o estudo dos reservatórios de petróleo e gás. Esse arcabouço metodológico viabilizou uma análise multifacetada dos dados, enriquecendo a interpretação e facilitando a identificação de campos análogos. Tais técnicas demonstraram que a combinação de múltiplos métodos pode potencializar o entendimento, especialmente quando confrontados com vastos conjuntos de dados.

Ao examinar o campo de Serra e seus campos análogos, torna-se evidente a complexidade intrincada subjacente à pesquisa em reservatórios petrolíferos. A superficialidade das análises iniciais pode mascarar divergências críticas que surgem quando características mais detalhadas, como o ambiente deposicional e a qualidade do reservatório, são consideradas. Os campos de Hawkings, Hastings e Loudon ilustram esta complexidade, com cada um deles apresentando peculiaridades que poderiam influenciar as táticas de exploração e os métodos de recuperação.

O alinhamento percebido nas variáveis categóricas entre os campos sugere que o campo de Serra está congruente com padrões convencionais observados nos campos análogos em termos de mecanismos de produção e fundamentos geológicos. No entanto, variações em aspectos como a qualidade do reservatório e a composição primária emergem como tópicos primordiais para futuras investigações. Estas distinções não apenas têm potencial para impactar a eficiência de métodos específicos de EOR, mas também podem recalibrar expectativas sobre a longevidade produtiva do reservatório.

Vale ressaltar que os campos análogos identificados atendem critérios globais e diretrizes estabelecidas por entidades como a ANP. Esta conformidade reforça a robustez da metodologia e a validade dos campos selecionados como análogos confiáveis. Além disso, a ferramenta de análise proposta é de imenso valor, especialmente para empresas emergentes no setor que buscam avaliar estratégias de recuperação sem a necessidade de softwares de simulação avançada. Em cenários de aquisições ou ofertas de BID, a capacidade de avaliar

rapidamente a produtividade potencial e identificar estratégias de recuperação otimizadas, com base em campos semelhantes, torna-se uma vantagem competitiva indispensável.

7. CONCLUSÕES

A rotina metodológica aqui desenvolvida e aplicada integra diversas modalidades de informação, processamento e visualização de dados, permitindo-se explorar um amplo conjunto de experiências de projetos anteriores de EOR em escala mundial. Tal abordagem permitiu obter uma representação de conhecimento autocontida de alto nível sobre as experiências de EOR, ter um panorama mundial da aplicação de EOR e compreender inter-relações implícitas entre variáveis com os diferentes métodos de EOR, possibilitando identificar tendências, novas direções de pesquisa e, assim, tomar decisões mais bem informadas.

Com base nos resultados obtidos, foi possível notar que variáveis de diferentes naturezas demonstram impacto na escolha, aplicação e sucesso técnico, econômico e tecnológico dos métodos de EOR. Tais variáveis indicam ter impactos que variam de acordo não apenas com as propriedades técnicas dos reservatórios e o método de EOR em questão, mas também com aspectos locais, comerciais e temporais. Essa análise permite esclarecer parte da complexa e intrincada rede de fatores que influenciam na aplicação de métodos de EOR em escala global. Tem como principal contribuição também a análise visual da interdependência de variáveis de diferentes naturezas e que demonstram ter impacto na escolha, aplicação e desenvolvimento de métodos de EOR.

Além disso, apesar da difusão de tecnologias de recuperação secundária e terciária, ainda se mostra concentração de diferentes métodos de EOR em algumas poucas regiões do mundo. Foram identificados fatores para a implementação, manutenção e expansão de projetos de EOR que vão desde razões técnicas, de existência de reservas depletadas, de infraestrutura disponível e fornecimento e disponibilidade de insumos, até aspectos relativos a

políticas de incentivos e aspectos regulatórios de cada país, estrutura e posicionamento estratégico dos segmentos de empresas e cenário econômico mundial.

Observou-se que fatores técnicos relacionados a propriedades litológicas e de fluxo dos reservatórios, como porosidade, permeabilidade, profundidade, temperatura, viscosidade, °API, salinidade da formação, heterogeneidades, dentre outros possuem diferentes impactos na escolha de métodos de EOR distintos. A criação de variáveis sintéticas permitiu identificar correlações e padrões antes implícitos e aprofundar a análise das variáveis técnicas e seus impactos nos métodos de EOR.

Foi observada também a influência que variáveis não petrofísicas possuem na aplicação de métodos de EOR. Aspectos relacionados à disponibilidade e custos de compra e transporte de fluidos para injeção demonstraram apresentar grande influência nos processos de tomada de decisão quanto à escolha do método de EOR a ser aplicado. Como exemplo, podemos destacar a injeção de gás obtido de campos vizinhos, a injeção de CO₂ com fornecimento provindo de rede captura de indústrias e gasodutos específicos, e a injeção de produtos químicos em mercados com indústria desenvolvida e de baixo custo.

Verificou-se influência do tipo de ambiente do campo em questão, se *offshore* ou *onshore*, em que viabilidade técnica e econômica se mostra ainda desafiadora. A viabilidade técnica/econômica mostrou poder ser prejudicadas devido a fatores como grandes profundidades e lâminas d'água, implicando em dificuldades como perdas de calor ou energia na injeção de fluidos no reservatório; maiores distâncias entre os poços, exigindo maior volume de produtos injetados e dificuldades; necessidade de adaptações no tratamento e separação do óleo; limitação de espaço e de capacidade de processamento, dentre outras. Nesse sentido, apesar de haver projetos em escala de campo para métodos de EOR já consolidados, verifica-se falta de experiência da indústria do petróleo na aplicação de métodos de EOR em ambiente *offshore*, sobretudo em lâmina d'água profunda, indicando sua implantação comercial ainda ser incipiente a projetos pontuais e um desafio com grande potencial a ser superado. Além disso, os projetos de EOR ainda representam uma parcela pequena da produção mundial, mas com grande potencial de crescimento. Historicamente,

observou-se que a adoção de EOR em larga escala tem se concentrado nos EUA, Canadá, China e Noruega – em grande parte associada a políticas de incentivo de cada um dos países. Porém, nos últimos anos outros países como Malásia, Emirados Árabes Unidos, Kuwait, Arábia Saudita, Oman, Índia, Colômbia e Equador também passaram a ter projetos piloto de EOR. Além disso, países como Trindade e Tobago, Índia e Indonésia vêm implementando dispositivos regulatórios específicos de incentivos fiscais à adoção de EOR.

Verificou-se que há uma dependência do preço do barril do petróleo para execução dos projetos de EOR. Além disso, também observa-se uma correlação do preço do barril com a pesquisa, aplicação, continuidade e avanço da aplicação de técnicas de EOR para novas fronteiras. Tal impacto do baixo preço do barril apresenta-se variar conforme o tipo de método de EOR, sendo mais significativo para tecnologias com menor grau de maturidade, como os métodos químicos.

Também verificou-se uma diversidade de posicionamentos estratégicos em relação à preferência de projetos e técnicas de EOR conforme o perfil de empresa operadora, havendo nichos específicos para cada tipo e porte de empresa operadora, com as majors (NOCs e IOCs) focando em ativos de grande classe e explorando o potencial de novas tecnologias e maior risco; e as pequenas e médias empresas com ênfase em ativos em sua maior parte já amplamente depletados e tecnologias de EOR já consolidadas. Nos últimos anos verifica-se que uma das tendências é fazer uso de processos combinados de EOR (ex: foam juntamente com injeção de gás; surfactante com injeção de vapor, etc), especialmente quando apenas um método de EOR não se encaixa perfeitamente nas especificações técnicas do reservatório. Porém, tais alternativas tornam mais complexo o processo de estudo, planejamento e a aplicação, sendo utilizado em geral apenas por NOCs e IOCs, e concentradas em sua maioria ainda em projetos pilotos.

Embora o custo do ciclo de vida do EOR possa ser competitivo com outras oportunidades de produção, ele frequentemente envolve altos requisitos de capital inicial e longos períodos de retorno. Como resultado, a produção de EOR historicamente dependeu de alguma forma de suporte ou escolha estratégica. Até o momento grande parte dos projetos de EOR implementados se beneficiou de algum tipo de incentivo governamental ou é iniciativa de NOCs

como parte de seus esforços para maximizar o retorno dos recursos nacionais e desenvolvimento de novas tecnologias. Em particular, a injeção de CO₂ para EOR associado com CCS tem sido vista como uma tecnologia capaz de reduzir a intensidade de emissões do petróleo entregue ao mercado e, portanto, reduzir sua pegada ambiental, tendo recebido interesse, apoio político e incentivos fiscais adicionais nos últimos anos.

Como boas práticas em geral, identificou-se a importância de incentivos fiscais e arcabouço regulatório adequado, do desenvolvimento de infraestrutura, da necessidade de planejamento da aplicação de EOR, incluindo o *screening*, aproveitar ao máximo o potencial da digitalização para gerar uma melhor compreensão do subsolo, a utilização de projetos e experiências prévias como base e estudos mais detalhados para redução de incerteza.

Os resultados da metodologia de obtenção de campos análogos para um dado reservatório de interesse a partir da base de dados consolidada mostrou constituir-se em um fluxo que permite a obtenção de campos análogos de maneira criteriosa, além de possibilitar repensar estratégias de produção, e avaliar métodos de produção implementados em outros campos semelhantes no mundo de uma forma rápida. Tal processo pode permitir uma aceleração do fluxo de planejamento da produção e comparação de desempenho e avaliação complementar de estratégias, reduzindo custos, tempos de simulação e análises de laboratório. Traz ainda relevante vantagem competitiva e celeridade para a elaboração de propostas de aquisição de novos ativos em diferentes contextos geológicos, além de poder atuar como uma ferramenta de suporte para planos de desenvolvimento e identificação de análogos atendendo critérios exigidos por agências reguladoras e diretrizes internacionais. Principais limitações da metodologia inclui o fato de características dos campos muitas vezes com formações amplamente heterogêneas, terem de ser resumidas por um único valor numérico, podendo não ser fidedigno da realidade geológica e de produção em questão, o que pode ser objeto de desenvolvimento de próximos trabalhos.

Com base nisso, conclui-se que tal trabalho permitiu a consolidação de uma ampla base de dados de projetos de EOR e de produção em escala mundial, desenvolveu análises elucidando aspectos diversos em relação à aplicação de EOR, e a criação de metodologia para obtenção de campos análogos validada por critérios técnicos. Tais análises realizadas e metodologias

desenvolvidas podem servir de base para subsidiar decisões mais bem informadas na indústria de óleo e gás, especialmente no contexto de desenvolvimento de campos maduros e métodos avançados de recuperação de petróleo.

8. REFERÊNCIAS BIBLIOGRÁFICAS

ABIDIN, A. Z.; PUSPASARI, T.; NUGROHO, W. A. Polymers for Enhanced Oil Recovery Technology. **Procedia Chemistry**, v. 4, p. 11–16, 2012.

AGGARWAL, C. C. **Data Mining**. Cham: Springer International Publishing, 2015.

ALADASANI, A.; BAI, B. **Recent Developments and Updated Screening Criteria of Enhanced Oil Recovery Techniques**. In: INTERNATIONAL OIL AND GAS CONFERENCE AND EXHIBITION IN CHINA. OnePetro, 8 jun. 2010. Disponível em: <<https://onepetro.org/SPEIOGCEC/proceedings/10IOGCEC/All-10IOGCEC/SPE-130726-MS/106815>>. Acesso em: 24 set. 2021

ALVARADO, V.; MANRIQUE, E. Enhanced Oil Recovery: An Update Review. **Energies**, v. 3, n. 9, p. 1529–1575, 27 ago. 2010a.

ALVARADO, V.; MANRIQUE, E. Enhanced Oil Recovery: An Update Review. **Energies**, v. 3, n. 9, p. 1529–1575, 27 ago. 2010b.

AWAN, A. R. A Survey of North Sea Enhanced-Oil-Recovery Projects Initiated During the Years 1975 to 2005. p. 17, 2008.

BABADAGLI, T. Development of mature oil fields — A review. **Journal of Petroleum Science and Engineering**, v. 57, n. 3–4, p. 221–246, jun. 2007.

BABADAGLI, T. Philosophy of EOR. **Journal of Petroleum Science and Engineering**, v. 188, p. 106930, maio 2020.

BELAZREG, L.; MAHMOOD, S. M. Water alternating gas incremental recovery factor prediction and WAG pilot lessons learned. **Journal of Petroleum Exploration and Production Technology**, v. 10, n. 2, p. 249–269, fev. 2020.

BELHAJ, A. F. et al. The effect of surfactant concentration, salinity, temperature, and pH on surfactant adsorption for chemical enhanced oil recovery: a review. **Journal of Petroleum Exploration and Production Technology**, v. 10, n. 1, p. 125–137, jan. 2020.

BERTINI, E.; TATU, A.; KEIM, D. Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2203–2212, dez. 2011.

BONDOR, P. L. Planning EOR Projects in *Offshore* Oil Fields. p. 7, 2005.

BRODIE, J. A. et al. **Review Of Gas Injection Projects In BP**. In: SPE IMPROVED OIL RECOVERY SYMPOSIUM. Society of Petroleum Engineers, 1 jan. 2012. Disponível em: <<https://www.onepetro.org/conference-paper/SPE-154008-MS>>. Acesso em: 22 nov. 2020

CHEN, J. et al. The potential of microalgae in biodiesel production. **Renewable and Sustainable Energy Reviews**, v. 90, p. 336–346, 1 jul. 2018.

CHEN, Q. et al. The prospects and potential opportunities of low salinity water flooding for *offshore* applications in sandstones. **Journal of Petroleum Science and Engineering**, v. 199, p. 108260, 1 abr. 2021.

CHOI, S. Microscale microbial fuel cells: Advances and challenges. **Biosensors and Bioelectronics**, v. 69, p. 8–25, 15 jul. 2015.

CHU, X. et al. **Data Cleaning: Overview and Emerging Challenges**. Proceedings of the 2016 International Conference on Management of Data. **Anais...** In: SIGMOD/PODS'16: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA. San Francisco California USA: ACM, 26 jun. 2016. Disponível em: <<https://dl.acm.org/doi/10.1145/2882903.2912574>>. Acesso em: 15 ago. 2021

CIPRIANI, G. et al. Perspective on hydrogen energy carrier and its automotive applications. **International Journal of Hydrogen Energy**, v. 39, n. 16, p. 8482–8494, 27 maio 2014.

DASAN, Y. K. et al. Life cycle evaluation of microalgae biofuels production: Effect of cultivation system on energy, carbon emission and cost balance analysis. **Science of The Total Environment**, v. 688, p. 112–128, 20 out. 2019.

DELAMAIDE, E.; DEGRE, G. Chemical EOR for Heavy Oil: the Canadian Experience. p. 31, [s.d.].

DIAMANTINI, C.; POTENA, D.; STORTI, E. A virtual mart for knowledge discovery in databases. **Information Systems Frontiers**, v. 15, n. 3, p. 447–463, jul. 2013.

DONG, X. et al. Enhanced oil recovery techniques for heavy oil and oilsands reservoirs after steam injection. **Applied Energy**, v. 239, p. 1190–1211, abr. 2019.

EIA. **Short-Term Energy Outlook - U.S. Energy Information Administration (EIA)**. Disponível em: <<https://www.eia.gov/outlooks/steo/report/prices.php>>. Acesso em: 29 set. 2020.

FAKHER, S.; IMQAM, A. A data analysis of immiscible carbon dioxide injection applications for enhanced oil recovery based on an updated database. **SN Applied Sciences**, v. 2, n. 3, p. 448, mar. 2020.

FAYYAD, U. 2. Knowledge Discovery In Databases: An Overview. v. 17, n. AI Magazine Volume, p. 20, 1996.

FORBES. **Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says**. Disponível em: <<https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>>. Acesso em: 28 set. 2021.

GANGULI, S.; SOMPOLINSKY, H. Compressed Sensing, Sparsity, and Dimensionality in Neuronal Information Processing and Data Analysis. **Annual Review of Neuroscience**, v. 35, n. 1, p. 485–508, 21 jul. 2012.

GARDNER, W. et al. Visualizing ToF-SIMS Hyperspectral Imaging Data Using Color-Tagged Toroidal Self-Organizing Maps. **Analytical Chemistry**, v. 91, n. 21, p. 13855–13865, 5 nov. 2019.

GBADAMOSI, A. O. et al. A review of gas enhanced oil recovery schemes used in the North Sea. **Journal of Petroleum Exploration and Production Technology**, v. 8, n. 4, p. 1373–1387, dez. 2018.

GBADAMOSI, A. O. et al. An overview of chemical enhanced oil recovery: recent advances and prospects. **International Nano Letters**, v. 9, n. 3, p. 171–202, set. 2019.

GEBREMARIAM, S. N.; MARCHETTI, J. M. Economics of biodiesel production: Review. **Energy Conversion and Management**, v. 168, p. 74–84, 15 jul. 2018.

GOH, B. H. H. et al. Sustainability of direct biodiesel synthesis from microalgae biomass: A critical review. **Renewable and Sustainable Energy Reviews**, v. 107, p. 59–74, 1 jun. 2019.

GONZÁLEZ-CALABOZO, J. M.; VALVERDE-ALBACETE, F. J.; PELÁEZ-MORENO, C. Interactive knowledge discovery and data mining on genomic expression data with numeric formal concept analysis. **BMC Bioinformatics**, v. 17, n. 1, p. 374, dez. 2016.

GREGORY PIATETSKY. **CRISP-DM, still the top methodology for analytics, data mining, or data science projects**.

HAMA, M. Q. et al. **Updated Screening Criteria for Steam Flooding Based on Oil Field Projects Data**. SPE Heavy Oil Conference-Canada. **Anais...** In: SPE HEAVY OIL CONFERENCE-CANADA. Calgary,

Alberta, Canada: Society of Petroleum Engineers, 2014. Disponível em: <<http://www.onepetro.org/doi/10.2118/170031-MS>>. Acesso em: 23 nov. 2020

HASHEMI-KIASARI, H. et al. Effect of operational parameters on SAGD performance in a dip heterogeneous fractured reservoir. **Fuel**, v. 122, p. 82–93, abr. 2014.

HIRVE, S.; PRADEEP REDDY, C. H. A Survey on Visualization Techniques Used for Big Data Analytics. In: BHATIA, S. K. et al. (Eds.). **Advances in Computer Communication and Computational Sciences**. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore, 2019. v. 924p. 447–459.

HUBER, S. et al. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. **Procedia CIRP**, v. 79, p. 403–408, 2019.

IEA. **OECD total primary energy supply (TPES) by source, 1971-2018 – Charts – Data & Statistics**. Disponível em: <<https://www.iea.org/data-and-statistics/charts/oecd-total-primary-energy-supply-tpes-by-source-1971-2018>>. Acesso em: 30 set. 2020.

ITOH, T. et al. High-dimensional data visualization by interactive construction of low-dimensional parallel coordinate plots. **Journal of Visual Languages & Computing**, v. 43, p. 1–13, dez. 2017.

JERAULD, G. R.; SALTER, S. J. The effect of pore-structure on hysteresis in relative permeability and capillary pressure: Pore-level modeling. **Transport in Porous Media**, v. 5, n. 2, p. 103–151, abr. 1990.

JI-HONG, H. et al. Production Performance of In-Situ Combustion in Heavy Oil Reservoir after Steam Injection. **Xinjiang Petroleum Geology**, v. 31, n. 5, p. 517, 21 out. 2020.

JIN, L. Advancing CO2 enhanced oil recovery and storage in unconventional oil play—Experimental studies on Bakken shales. p. 13, 2017.

KADI, I.; IDRI, A.; FERNANDEZ-ALEMAN, J. L. Knowledge discovery in cardiology: A systematic literature review. **International Journal of Medical Informatics**, v. 97, p. 12–32, jan. 2017a.

KADI, I.; IDRI, A.; FERNANDEZ-ALEMAN, J. L. Knowledge discovery in cardiology: A systematic literature review. **International Journal of Medical Informatics**, v. 97, p. 12–32, jan. 2017b.

KANG, P.-S.; LIM, J.-S.; HUH, C. *Screening Criteria and Considerations of Offshore Enhanced Oil Recovery*. **Energies**, v. 9, n. 1, p. 44, jan. 2016.

KE, C.-Y. et al. A pilot study on large-scale microbial enhanced oil recovery (MEOR) in Baolige Oilfield. **International Biodeterioration & Biodegradation**, v. 127, p. 247–253, 1 fev. 2018.

KOHONEN, T. Learning Vector Quantization. In: KOHONEN, T. (Ed.). **Self-Organizing Maps**. Springer Series in Information Sciences. Berlin, Heidelberg: Springer, 2001. p. 245–261.

KOHONEN, T. Essentials of the self-organizing map. **Neural Networks**, v. 37, p. 52–65, jan. 2013.

KOKAL, S.; AL-KAABI, A. Enhanced oil recovery: challenges & opportunities. **World Petroleum Council: Official Publication 2010**, p. 6, 2010.

KOLODZIE, S. **Analysis Of Pore Throat Size And Use Of The Waxman-Smits Equation To Determine Ooip In Spindle Field, Colorado**. In: SPE ANNUAL TECHNICAL CONFERENCE AND EXHIBITION. OnePetro, 21 set. 1980. Disponível em: <<https://onepetro.org/SPEATCE/proceedings/80SPE/All-80SPE/SPE-9382-MS/111215>>. Acesso em: 27 set. 2021

KOWALEWSKI, E. et al. Microbial improved oil recovery—bacterial induced wettability and interfacial tension effects on oil production. **Journal of Petroleum Science and Engineering**, Reservoir Wettability. v. 52, n. 1, p. 275–286, 1 jun. 2006.

KUMAR, V. K.; FASSIHI, M. R. Case History and Appraisal of the Medicine Pole Hills Unit Air Injection Project. **SPE Reservoir Engineering**, v. 10, n. 03, p. 198–202, 1 ago. 1995.

LATIL, M. et al. **ENHANCED OIL RECOVERY**. [s.l: s.n.].

Custodio, L. F. N., & de Carvalho Carneiro, C. (2024). **Global enhanced oil recovery assessment: A comprehensive data-driven analysis using self-organizing maps**. *Geoenergy Science and Engineering*, 237, 212771. <https://doi.org/10.1016/j.geoen.2024.212771>

Custódio, L. F. N., Jeiss, B. W., Stefanelli, D., Kirch, A., Macedo, M., & de Carvalho Carneiro, C. (2024). **Obtenção de reservatórios análogos com o uso de machine learning em base de dados mundial de experiências de produção de campos de petróleo**. Apresentado no III Seminário de Inteligência Artificial Aplicada à Indústria do Petróleo (ABGP), Rio de Janeiro, Brasil.

LI, P. et al. *Screening* and simulation of offshore CO₂-EOR and storage: A case study for T the HZ21-1 oilfield in the Pearl River Mouth Basin, Northern South China Sea. **International Journal of Greenhouse Gas Control**, p. 16, 2019.

LI, Y.; HOU, S. **Methods and Techniques in Data Visualization Model**. 2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC). **Anais...** In: 2017 INTERNATIONAL CONFERENCE ON COMPUTER TECHNOLOGY, ELECTRONICS AND COMMUNICATION (ICCTEC). Dalian, China: IEEE, dez. 2017. Disponível em: <<https://ieeexplore.ieee.org/document/8789106/>>. Acesso em: 5 set. 2020

LIU, S. et al. Visualizing High-Dimensional Data: Advances in the Past Decade. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 3, p. 1249–1268, 1 mar. 2017.

LIU, Z. et al. Status and progress of worldwide EOR field applications. **Journal of Petroleum Science and Engineering**, v. 193, p. 107449, out. 2020.

LOTFNEZHAD AFSHAR, H. et al. Prediction of Breast Cancer Survival Through Knowledge Discovery in Databases. **Global Journal of Health Science**, v. 7, n. 4, p. p392, 26 jan. 2015.

LTIFI, H. et al. A human-centred design approach for developing dynamic decision support system based on knowledge discovery in databases. **Journal of Decision Systems**, v. 22, n. 2, p. 69–96, abr. 2013.

LYND, L. R. The grand challenge of cellulosic biofuels. **Nature Biotechnology**, v. 35, n. 10, p. 912–915, out. 2017.

MAHLIA, T. M. I. et al. Patent landscape review on biodiesel production: Technology updates. **Renewable and Sustainable Energy Reviews**, v. 118, p. 109526, 1 fev. 2020.

MANRIQUE, E. Effective EOR Decision Strategies With Limited Data: Field Cases Demonstration. p. 11, 2009.

MANRIQUE, E.; GURFINKEL, M.; MUCI, V. Enhanced Oil Recovery Field Experiences in Carbonate Reservoirs in the United States. p. 33, 2014.

MANRIQUE, E. J.; ENGINEERING, N. Q.; MUCI, V. E. EOR Field Experiences in Carbonate Reservoirs in the United States. p. 20, 2007.

MARTINEZ-PLUMED, F. et al. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. **IEEE Transactions on Knowledge and Data Engineering**, p. 1–1, 2020.

MASHAYEKHIZADEH, V.; KORD, S.; DEJAM, M. EOR POTENTIAL WITHIN IRAN. **Special Topics Reviews in Porous Media: An International Journal**, v. 5, n. 4, p. 325–354, 2014.

MCCOLLUM, D. L. et al. Quantifying uncertainties influencing the long-term impacts of oil prices on energy markets and carbon emissions. **Nature Energy**, v. 1, n. 7, p. 16077, jul. 2016.

MCKINSEY. **Oil and gas after COVID-19: The day of reckoning or a new age of opportunity? | McKinsey**. Disponível em: <<https://www.mckinsey.com/industries/oil-and-gas/our-insights/oil-and-gas-after-covid-19-the-day-of-reckoning-or-a-new-age-of-opportunity>>. Acesso em: 30 set. 2020.

MIKAEIL, R. Rock Penetrability Classification Using Artificial Bee Colony (ABC) Algorithm and Self-Organizing Map. **Geotech Geol Eng**, p. 10, 2018.

MILLER, R. G.; SORRELL, S. R. The future of oil supply. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 372, n. 2006, p. 20130179, 13 jan. 2014.

MILLER, R. J. **Koch's experience with deep in situ combustion in Williston basin**. [s.l.] BDM Oklahoma, Inc., Bartlesville, OK (United States); National Inst. for Petroleum and Energy Research, Bartlesville, OK (United States), 1 fev. 1995. Disponível em: <<https://www.osti.gov/biblio/79268>>. Acesso em: 23 set. 2021.

MOHD AZHAR, S. H. et al. Yeasts in sustainable bioethanol production: A review. **Biochemistry and Biophysics Reports**, v. 10, p. 52–61, 1 jul. 2017.

MOKHEIMER, E. M. A. et al. A Comprehensive Review of Thermal Enhanced Oil Recovery: Techniques Evaluation. **Journal of Energy Resources Technology**, v. 141, n. 3, p. 030801, 1 mar. 2019.

MUGGERIDGE, A. et al. Recovery rates, enhanced oil recovery and technological limits. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 372, n. 2006, p. 20120320, 13 jan. 2014.

MURER, A. S. et al. **Steam Injection Project In Heavy Oil Diatomite**. . In: SPE WESTERN REGIONAL MEETING. OnePetro, 25 jun. 1997. Disponível em: <<https://onepetro.org/SPEWRM/proceedings/97WRM/All-97WRM/SPE-38302-MS/188733>>. Acesso em: 22 set. 2021

NAGASHIMA, H.; KATO, Y. **APREP-DM: a Framework for Automating the Pre-Processing of a Sensor Data Analysis based on CRISP-DM**. 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). **Anais...** In: 2019 IEEE INTERNATIONAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS WORKSHOPS (PERCOM WORKSHOPS). Kyoto, Japan: IEEE, mar. 2019. Disponível em: <<https://ieeexplore.ieee.org/document/8730785/>>. Acesso em: 15 ago. 2021

NAKAGAWA, K. Analysis of earthquake-induced groundwater level change using self-organizing maps. **Environmental Earth Sciences**, p. 11, 2019.

NINGNING; WANG; JIANJUN. Characteristics of oil and gas discoveries in recent 20 years and future exploration in the world. **China Petroleum Exploration**, v. 23, n. 1, p. 44–53, 4 jan. 2018.

OLAJIRE, A. A. Review of ASP EOR (alkaline surfactant polymer enhanced oil recovery) technology in the petroleum industry: Prospects and challenges. **Energy**, v. 77, p. 963–982, dez. 2014.

PATEL, J. et al. Recent developments in microbial enhanced oil recovery. **Renewable and Sustainable Energy Reviews**, v. 52, p. 1539–1558, 1 dez. 2015.

PUSCH, G. Perspectives of European Improved Oil Recovery. **RESERVOIR ENGINEERING**, p. 8, [s.d.].

QIAN, J. Introducing self-organized maps (SOM) as a visualization tool for materials research and education. p. 14, 2019.

QIN, X. Making data visualization more efficient and effective: a survey. p. 25, 2019.

RAGHAV, R. S. et al. **A survey of data visualization tools for analyzing large volume of data in big data platform**. 2016 International Conference on Communication and Electronics Systems (ICCES). **Anais...** In: 2016 INTERNATIONAL CONFERENCE ON COMMUNICATION AND ELECTRONICS SYSTEMS (ICCES). Coimbatore, India: IEEE, out. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7889976/>>. Acesso em: 5 set. 2020

RAMOS, N. M. M. et al. Knowledge discovery of indoor environment patterns in mild climate countries based on data mining applied to in-situ measurements. **Sustainable Cities and Society**, v. 30, p. 37–48, abr. 2017.

RANEY, K. et al. Surface and Subsurface Requirements for Successful Implementation of *Offshore* Chemical Enhanced Oil Recovery. **SPE Production & Operations**, v. 27, n. 03, p. 294–305, 1 ago. 2012.

RASSENFOSS, S. From Bacteria to Barrels: Microbiology Having an Impact on Oil Fields. **Journal of Petroleum Technology**, v. 63, n. 11, p. 32–38, 1 nov. 2011.

REKSIDLER, R. et al. **Offshore Chemical Enhanced Oil Recovery**. OTC Brasil. **Anais...** In: OTC BRASIL. Rio de Janeiro, Brazil: *Offshore* Technology Conference, 2015. Disponível em: <<http://www.onepetro.org/doi/10.4043/26123-MS>>. Acesso em: 1 set. 2020

REZAEI, M. R.; JAFARI, A.; KAZEMZADEH, E. Relationships between permeability, porosity and pore throat size in carbonate rocks using regression analysis and neural networks. **Journal of Geophysics and Engineering**, v. 3, n. 4, p. 370–376, 1 dez. 2006.

ROGALEWICZ, M.; SIKI, R. Methodologies of Knowledge Discovery from Data and Data Mining Methods in Mechanical Engineering. **Management and Production Engineering Review**, v. 7, n. 4, p. 97–108, 1 dez. 2016.

SALEH, L. D.; WEI, M.; BAI, B. Data Analysis and Updated *Screening* Criteria for Polymer Flooding Based on Oilfield Data. **SPE Reservoir Evaluation & Engineering**, v. 17, n. 01, p. 15–25, 1 fev. 2014.

SEGEV, A. Identification of trends from patents using self-organizing maps. **Expert Systems with Applications**, p. 8, 2012.

SEGEV, A.; KANTOLA, J. Identification of trends from patents using self-organizing maps. **Expert Systems with Applications**, v. 39, n. 18, p. 13235–13242, dez. 2012.

SHAFIQUE, U.; QAISER, H. A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). v. 12, n. 1, p. 7, 2014.

SHENG, J. **Modern Chemical Enhanced Oil Recovery: Theory and Practice**. [s.l.] Gulf Professional Publishing, 2010.

SHENG, J. **Enhanced Oil Recovery Field Case Studies**. [s.l.] Gulf Professional Publishing, 2013.

SKIENA, S. S. **The Data Science Design Manual**. Cham: Springer International Publishing, 2017.

STEFANOVIČ, P.; KURASOVA, O. Visual analysis of self-organizing maps. **Nonlinear Analysis: Modelling and Control**, v. 16, n. 4, p. 488–504, 7 dez. 2011.

STUDER, S. et al. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. **arXiv:2003.05155 [cs, stat]**, 24 fev. 2021.

TABER, J. J.; MARTIN, F. D.; SERIGHT, R. S. EOR *Screening* Criteria Revisited - Part 1: Introduction to *Screening* Criteria and Enhanced Recovery Field Projects. **SPE Reservoir Engineering**, v. 12, n. 03, p. 189–198, 1 ago. 1997.

THOMAS, S. Enhanced Oil Recovery - An Overview. **Oil & Gas Science and Technology - Revue de l'IFP**, v. 63, n. 1, p. 9–19, jan. 2008.

TRIPATHI, S. et al. Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing. **arXiv:2007.14791 [cs]**, 28 jul. 2020.

TRUJILLO, M. et al. **Selection Methodology for Screening Evaluation of Enhanced-Oil-Recovery Methods**. In: SPE LATIN AMERICAN AND CARIBBEAN PETROLEUM ENGINEERING CONFERENCE. OnePetro, 1 dez. 2010. Disponível em: <<https://onepetro.org/SPELACP/proceedings/10LACP/All-10LACP/SPE-139222-MS/107368>>. Acesso em: 24 set. 2021

TURTA, A. T. et al. Current Status of Commercial In Situ Combustion Projects Worldwide. **Journal of Canadian Petroleum Technology**, v. 46, n. 11, p. 7, 2007.

WEVER, D. A. Z.; PICCHIONI, F.; BROEKHUIS, A. A. Polymers for enhanced oil recovery: A paradigm for structure–property relationship in aqueous solution. **Progress in Polymer Science**, v. 36, n. 11, p. 1558–1628, nov. 2011.

WIEMER, H.; DROWATZKY, L.; IHLENFELDT, S. Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. **Applied Sciences**, v. 9, n. 12, p. 2407, 13 jun. 2019.

WIRTH, R.; HIPPEL, J. CRISP-DM: Towards a Standard Process Model for Data Mining. p. 11, 12 jun. 2000.

YANG, H.-C.; LEE, C.-H.; WU, C.-Y. Sentiment Discovery of Social Messages Using Self-Organizing Maps. **Cognitive Computation**, v. 10, n. 6, p. 1152–1166, dez. 2018.

YOKOTA, H. et al. Framework for visual-feedback training based on a modified self-organizing map to imitate complex motion. **Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology**, v. 234, n. 1, p. 49–58, mar. 2020.

YUAN, B.; WOOD, D. A. A comprehensive review of formation damage during enhanced oil recovery. **Journal of Petroleum Science and Engineering**, v. 167, p. 287–299, ago. 2018.

YUAN, S.; WANG, Q. New progress and prospect of oilfields development technologies in China. **Petroleum Exploration and Development**, v. 45, n. 4, p. 698–711, ago. 2018.

ZABED, H. et al. Bioethanol production from renewable sources: Current perspectives and technological progress. **Renewable and Sustainable Energy Reviews**, v. 71, p. 475–501, 1 maio 2017.

ZHANG, N.; WEI, M.; BAI, B. Statistical and analytical review of worldwide CO₂ immiscible field applications. **Fuel**, v. 220, p. 89–100, maio 2018.

9. ANEXOS

Anexo A – Trabalho apresentado no III Seminário de Inteligência Artificial Aplicada à Indústria de Petróleo (ABGP)

OBTENÇÃO DE RESERVATÓRIOS ANÁLOGOS COM O USO DE MACHINE LEARNING EM BASE DE DADOS MUNDIAL DE EXPERIÊNCIAS DE PRODUÇÃO DE CAMPOS DE PETRÓLEO

Luiz Felipe Niedermaier Custódio^{1,2}, Bruno Wamzer Jeiss^{2,3}, Denilson Stefanelli^{2,3}, Alexsandro Kirch^{2,4}, Marcellus Macedo², Cleyton de Carvalho Carneiro¹

¹Departamento de Engenharia Naval e Oceânica da Poli-USP (PPGEN), Av. Professor Mello Moraes, 2231 - Butantã, São Paulo - SP, 05508-030 (luiz.custodio@usp.br);

²DEEPSTRATA, Avenida Paulista, 1636, Bela Vista, São Paulo – SP 01310-200;

³Faculdade de Engenharia Mecânica - Unicamp, Rua Mendeleev, 200 - Campinas, São Paulo - CEP 13083-860;

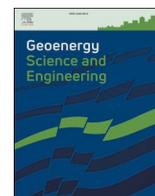
⁴Departamento de Engenharia Mecânica da Poli-USP, Av. Professor Mello Moraes, 2231 - Butantã, São Paulo - SP, 05508-030;

Reservatórios análogos, segundo o *Petroleum Resources Management System*, são aqueles com propriedades semelhantes de rocha e fluido, condições de reservatório (profundidade, temperatura e pressão) e de mecanismos de produção, e em um estágio mais avançado de desenvolvimento em relação ao reservatório de interesse. Sua identificação pode auxiliar na interpretação de dados nos planejamentos de exploração e desenvolvimento, na avaliação de estratégias de produção e estimativas de recuperação. Contudo, encontrar os análogos de um reservatório consiste em tarefa complexa, uma vez que envolve a correspondência de variáveis diversas, incluindo propriedades geológicas, petrofísicas, dos fluidos e de engenharia. As rotinas existentes para identificação de campos análogos dependem do julgamento do avaliador e da sua experiência em vários domínios das geociências e engenharias. Comumente as rotinas fazem uso de um único análogo, normalmente de formações geológicas vizinhas ao reservatório de interesse, o que acarreta riscos na caracterização de uma oportunidade e um espectro restrito e enviesado de avaliação e obtenção de *insights*. A pesquisa aqui apresentada sumariza um procedimento sistemático para a busca de reservatórios análogos com base em um banco de dados contendo parâmetros geológicos e de engenharia de campos de petróleo ao redor do mundo. Foi desenvolvido um estudo de caso objetivando a obtenção de reservatórios análogos ao de um campo da Bacia Potiguar, tendo como vista critérios estabelecidos pela ANP. O método utiliza técnicas estatísticas multivariadas e de aprendizado de máquina para encontrar reservatórios que sejam mais semelhantes ao campo de interesse. A base de dados construída possui 1735 projetos de produção em campos de petróleo localizados em 29 países e que englobam 112 atributos técnicos. A compilação dos dados foi feita a partir de artigos científicos, relatórios de órgãos especializados, dados de agências reguladoras de cada país, apresentações para investidores de companhias operadoras, dentre outras fontes. A base de dados resultante contém sete principais conjuntos de variáveis para cada reservatório: (i) registros de localização; (ii) características da bacia, da formação produtora e do contexto geológico; (iii) propriedades petrofísicas básicas e avançadas; (iv) propriedades do óleo; (v) propriedades termodinâmicas; (vi) empresas operadoras, números de poços e variáveis de projeto e (vii) histórico de produção. O protocolo metodológico consiste em quatro etapas: (a) pré-processamento, com seleção de parâmetros-chave; (b) análise multivariada; (c) aplicação de algoritmos de clusterização hierárquica e ranqueamento por medidas de similaridade e (d) análise qualitativa e detalhada dos candidatos análogos em maior potencial. Esta abordagem demonstrou flexibilidade para avaliar diferentes cenários e prioridades em relação às variáveis de maior interesse, além de viabilizar comparações e aprendizado com experiências internacionais. Tal sistematização da busca por reservatórios análogos, usados em conjunto com os dados primários, expandem o conhecimento do indivíduo e da equipe. Os resultados desenvolvem percepções que não seriam viáveis com o uso de dados locais ou da experiência individual isoladamente, representando um ganho na assertividade em tomadas de decisão relativas à E&P.

PALAVRAS-CHAVE: CAMPOS ANÁLOGOS, BASE DE DADOS MUNDIAL, MACHINE LEARNING

APOIO: Os autores agradecem à Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) pela concessão de bolsa de mestrado ao primeiro autor, ao *Integrated Technology for Rock and Fluid Analysis* (InTRA) e ao Programa de Pós-Graduação do Departamento de Engenharia Naval e Oceânica da Escola Politécnica da Universidade de São Paulo (PPGEN).

Anexo B – Artigo publicado na revista *Geoenergy Science and Engineering*



Global enhanced oil recovery assessment: A comprehensive data-driven analysis using self-organizing maps

Luiz Felipe Niedermaier Custodio^{a,*}, Cleyton de Carvalho Carneiro^b

^a Universidade de São Paulo, USP, Escola Politécnica, Department of Naval and Oceanic Engineering, Av. Professor Mello Moraes, 2231, Butantã, São Paulo, SP, 05508-030, Brazil

^b Universidade de São Paulo, USP, Escola Politécnica, Department of Mining and Petroleum Engineering, Praça Narciso de Andrade, Vila Mathias, Santos, São Paulo, 11013-560, Brazil

ARTICLE INFO

Keywords:

Global EOR analysis
Oil reservoirs projects comparison
Data science
Multivariate data analysis
Machine learning

ABSTRACT

This study presents a comprehensive analysis of Enhanced Oil Recovery (EOR) methods' application in the global oil industry, identifying trends and main factors influencing their adoption. A wealth of unstructured and scattered data from public documents covering traditional EOR methods and newer technologies was processed to create a reliable global database, encompassing 1237 EOR projects in reservoirs around the world. Data science techniques and self-organizing maps (SOM) were employed for data analysis, dimensionality reduction, and visualization of complex problems, proving suitable for high-dimensionality and incomplete data. The research integrates various data modalities and data science visualization techniques, enabling a high-level representation of EOR projects' self-contained knowledge, and providing a comprehensive view of EOR application worldwide. The study reveals implicit interrelationships between multiple variables and EOR methods, identifying trends, research and development directions, and better-informed decisions. In-depth analysis showed that factors impacting the choice, application, and success of EOR methods include technical properties of reservoirs, local, commercial, and temporal aspects. Despite the diffusion of secondary and tertiary recovery technologies, a concentration of EOR methods exists in specific regions. Factors influencing EOR projects' implementation, maintenance, and expansion range from technical reasons, deselected reserves, and infrastructure to incentive policies and regulatory aspects of each country, business structure, and global economic scenario. Both petrophysical and non-petrophysical variables influence the application of EOR methods. Availability and costs of purchasing and transporting injection fluids greatly impact decision-making processes. The analysis contributes to understanding the complex network of factors influencing EOR, and this work can aid in making more informed decisions in the industry and research for EOR applications.

1. Introduction

The Enhanced Oil Recovery (EOR) methods, also called tertiary recovery, consist of the extraction of oil from a field from the injection of fluids. The injected fluids, in addition to providing extra energy to the reservoir, can also modify the properties of oil in the rock-fluid interaction to extract it more efficiently (Abidin et al., 2012; Wever et al., 2011). The application of these methods prolongs the useful life of oil fields and promotes a substantial increase in production. The technological improvement related to the EOR methods is an effective way to increase recovery factors, optimizing the production of fields in the various stages of production, especially in mature fields.

There are four main classes of EOR methods: (1) thermal methods;

(2) gas injection; (3) injection of chemicals; and (4) others. These classes can be subdivided into other different associated EOR methods, as shown in Fig. 1, which is adapted from. Each of the EOR methods tends to be applied to certain types of reservoirs, without a method considered globally the best. The success of the implementation of these methods depends both on the various properties of rocks and fluids of the reservoir, as well as on the structural model, and on the geographic, logistical, and economic factors, among others. In addition, there may be a combination of one or more EOR methods in the application in an oil field. This fact generates several possibilities of combining methods, which can make the evaluation and selection of EOR methods an even more complex task.

If, on the one hand, the evaluation and selection of methods is a complex activity, on the other hand, the amount and variety of data

* Corresponding author.

E-mail addresses: luiz.custodio@usp.br (L.F. Niedermaier Custodio), cleytoncarneiro@usp.br (C.C. Carneiro).

<https://doi.org/10.1016/j.geoen.2024.212771>

Received 30 September 2023; Received in revised form 12 February 2024; Accepted 13 March 2024

Available online 16 March 2024

2949-8910/© 2024 Published by Elsevier B.V.

List of abbreviations and acronyms	
BMU	Best Matching Unit
CCS	Carbon, Capture and Storage
CRISP	DM Cross-Industry Standard Process for Data Mining
EOR	Enhanced Oil Recovery,
HC	Hydrocarbon,
HWF	Hot Water Flooding
IOC	International Oil Company
ISC	In-Situ Combustion
KDD	Knowledge Discovery in Databases
LSWF	Low Salinity Waterflooding
NOC	National Oil Company
SAGD	Steam-Assisted Gravity Drainage
SOMs	Self-Organizing Maps
SP	Surfactant and Polymer
WAG	Water Alternating Gas

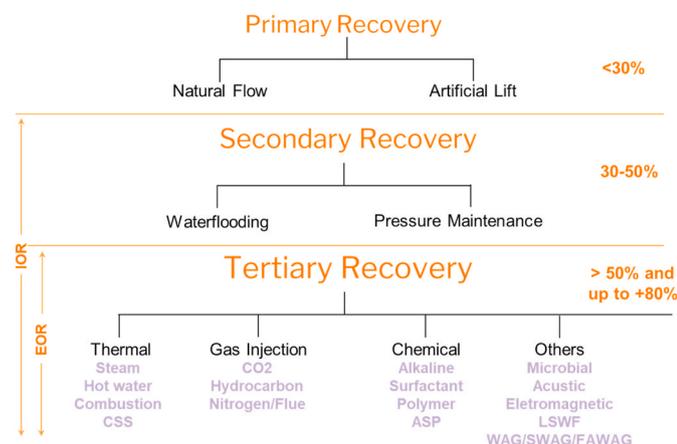


Fig. 1. Different oil recovery methods and their respective subclasses are commonly used in the development phases of an oil field. Adapted from [Kokal and Al-Kaabi \(2010\)](#).

generated and made publicly available increase exponentially. There is also a growing number of scientific articles and reports prepared by consultancies and official administrative bodies from various countries focused on energy security that address the application of EOR methods, including works that report case studies regarding the application of EOR at pilot scale, reservoir, or field level in various locations around the world ([Babadagli, 2007](#); [Babadagli, 2020](#); [Fakher and Imqam, 2020](#); [Manrique, 2009](#); [Muggeridge et al., 2014](#); [Olajire, 2014](#); [Thomas, 2008](#); [Zhang et al., 2018](#)). [Fig. 2](#) shows the number of publications present in the *OnePetro* database, which brings together publications on oil from around the world, relating to each EOR category over time. In particular, it is observed that even with the reduction in the price of a barrel of oil in 2014, the number of scientific articles addressing the EOR methods has grown significantly, constituting one of the possible sets of sources to obtain learning about previous experiences of EOR.

Several public documents often contain data, although sparse and in an isolated and unstructured way, about the application of EOR methods in oil fields worldwide. Such data sources, in addition to gathering information on more traditional fields and methods of EOR, such as steam injection and hydrocarbon injection in conventional reservoirs, sometimes also provide information on new technologies, such as *Low Salinity Waterflooding* (LSWF). Although in different formats, quite fragmented, and dispersed, these different data sources, with proper treatment for

information capture and processing, have ample informative potential. This information gathered made it possible to build a comprehensive and reliable global database with relevant information on EOR, capable of promoting targeted studies using advanced analysis techniques.

The self-organizing maps (SOM) technique consists of an ordered vector quantization tool developed in the n -dimensional space of the variables, which provides measures of similarity between the instances involved. The technique is internationally used in several applications, such as dimensionality reduction, analysis, and visualization of complex problems. SOM analyses, in addition to being suitable for high-dimensionality data, can also be applied to databases with incomplete or partially missing data. Additionally, it stands out from other unsupervised learning techniques by allowing the preservation of the topological structure and neighborhood among the data, as well as enabling the visualization of clusters influenced by each of the variables of interest.

This research aims to analyze the application of EOR in the world based on real projects of the implementation of EOR on a global scale. The projects were integrated into a single database, processed by SOM, and interpreted with the aid of data science tools. From the generated maps, it is intended (i) to develop a comprehensive and worldwide analysis of the application of EOR methods; (ii) to discuss the interrelationship between variables with an emphasis on observed patterns; and (iii) to identify the main factors that influence the application of EOR methods.

2. Conceptual background

Self-organizing maps

Within the scope of artificial neural networks, SOM represents a bioinspired tool for data analysis and visualization with unsupervised learning. The SOM technique is capable of preserving topological relationships projected on a map that is based on the use of vector quantization developed in n -dimensional space as well as on measurements of the similarity vector. SOMs computationally mimic functions similar to biological neural networks, such as the internal structure of input signals, the coordination of unit activities through lateral connections between units, and the specialization of certain sets of neurons with certain data clusters ([Kohonen, 2013](#)).

The SOM technique can be seen as an elastic artificial neural tissue of neurons adjusted to the input signal space, preserving the topological relationships of the signal space ([Kohonen, 2001](#)). [Table 3](#) summarizes the basic principles involved in the SOM algorithm. The input signal space is the set of N -dimensional vectors $x_i = [x_{i1}, x_{i2}, \dots, x_{in}, \dots, x_{iN}]$ $i = 1, 2, 3, \dots, M$ for which they represent the X integrated database. Each vector will be treated as an EOR project applied in one *reservoir* $_i$ of the database, with its characteristics represented by the variables. The output layer is an ordered collection of neurons that are typically organized as hexagonal or rectangular networks. Each neuron is connected to the input layer through weight vectors, thus having two layers. The dimensionality of the weight vector of each output neuron is the same as the dimensionality of the input vFactor; therefore, $w_i = [w_{j1}, w_{j2}, \dots, w_{jn}, \dots, w_{jN}]$ for $j = 1, 2, 3, \dots, J$, where J is the total number of neurons on the map. These weight vectors of the output neurons are compared with the input vectors according to some distance measurements to determine the degree of activation of that neuron. The Euclidean distance is generally used as a criterion for comparing the input vector with the weight vectors of the output neurons. After comparison, the neuron whose reference vector has greater similarity with the cluster of reservoirs is identified and defined as the winning neuron, $b_i = \underset{j}{\operatorname{argmin}} \|x_i - w_j\|$ to $i = 1, 2, 3, \dots, M$ and $j = 1, 2, 3, \dots, J$, which is called the Best Matching Unit (BMU) for input (see [Table 3](#)).

After selecting the winning neuron b_i , the weight vector of the winning neuron is updated to make it closer to the input vector. In addition,

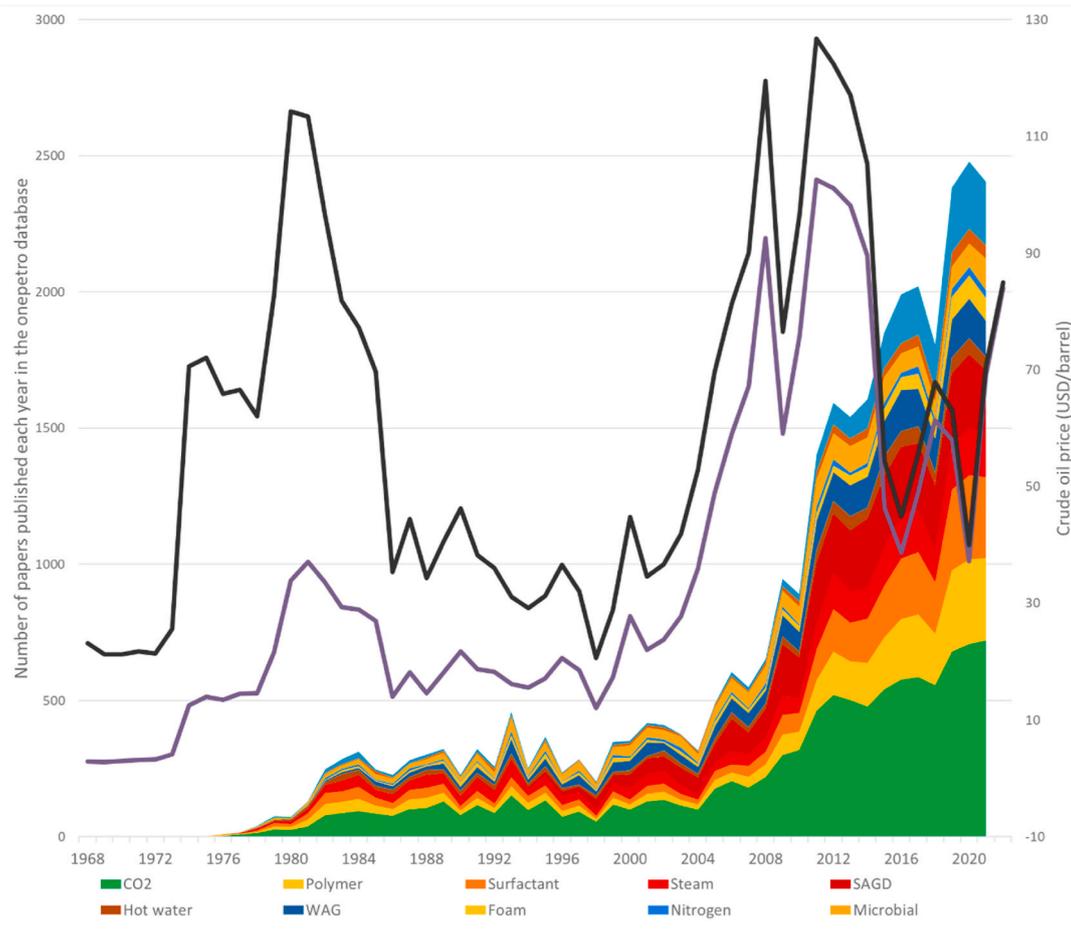


Fig. 2. Over the years on the OnePetro basis for each category of EOR and Brent oil barrel prices (nominal and debased inflation based on the US). The oil barrel price figures come from a report by the U.S. Energy Information Administration (EIA). The price discounted as U.S. inflation was also included to achieve a more balanced comparison in different economic scenarios.

Table 1

Source categories used to build the EOR reservoirs projects database.

Source Category	Description
Scientific Articles	Articles covering state-of-the-art EOR applications and field experiments in various production fields.
Comprehensive Surveys	Studies focusing on EOR methods in specific regions or sedimentary basins.
Field Experiments	Research focused on specific EOR methods such as CO2 injection, water-gas injection, microbial EOR, etc.
General Information Articles	Articles with broad, albeit less structured, EOR-related information.
International Reports & Databases	Publications from international agencies and institutes like IEA, CSIS, CCSI, OPEC, etc.
Regulatory and Research Bodies	Information from various national regulatory and research organizations.
Industry Reports & News	Reports from operating companies, investor presentations, and specialized news sources.
Books and Case Studies	Publications containing detailed case studies and examples of EOR applications.

the weight vectors of neighboring neurons are also updated, but to a lesser extent. The learning rate parameter and the neighborhood influence function can be defined as gradually decreasing during the training process. In this work, as a function of the neighborhood, the Gaussian function is given by:

$$h_{bj} = \exp\left(-\frac{\|r_b - r_j\|^2}{2\sigma^2(t)}\right)$$

where r_i is the position of neurons on the hexagonal map of the output neurons, r_j is the location of the j -th neuron, with $r_j \in \mathcal{R}^2$ being the position of the winning neuron on the output map, and $\sigma(t)$ is the neighborhood range of the influence of neurons, which reduces monotonically over time to ensure convergence. Such a neighborhood function is incorporated into the weight update function given by:

$$w_j(t+1) = w_j(t) + \alpha(t)h_{bj}(t) \|x_i(t) - w_j(t)\|$$

The adaptive process modifies the weights of the network as the network self-organization is achieved. The training phase ends when the number of times exceeds the predetermined total number of iterations, and the neurons on the output map will be labeled with their corresponding input names.

Two parameters to be specified that influence the accuracy of SOM are: (i) the number of neurons in the neural tissue and (ii) the dimensions of the height and width of the map. As the map size increases, the quantization error becomes smaller, but the topographic error, which is related to the projection of BMUs, becomes greater and the computational cost is higher. Sometimes, the dataset may contain only a few sets of patterns of interest to note, where a lower resolution is sufficient. However, if there is a suspicion that there are interesting structures in the data, a larger artificial neural tissue is necessary to obtain sufficient resolution for investigation. Typical SOM matrices range from a few tens to a few hundred nodes, and it is not possible to guess or estimate the most appropriate exact size of the matrix beforehand. It must be determined by the trial and error method, with quality checking to obtain a reasonable optimal solution. A compromise must be made

Table 2
Synthesis of the general structure of the database and variables collected about oil reservoirs in which EOR projects were applied.

Sets of variables present in the database			
Location records	Field name	Production data	Years of production
	Name of geological formation		Primary production mechanism
	Country		Time of the fields to reach the production plateau (years)
	State/province		Time of the fields in production on the plateau (years)
Basic reservoir properties	Administrative region	Project variables	Current recovery factor (%)
	County/county		Watercut (%)
	Latitude and longitude (°)		EUR (%)
	Dominant lithology		Design scale
Geological context of the reservoir	Porosity (%)	Recovery strategies	Project start and end date
	Permeability (mD)		Operating company
	Depth (ft)		Offshore environment
	Temperature (°F)		Water slide (ft)
Oil Properties	Net pay (ft)	Recovery efficiency	IOR/EOR method:
	Type of bowl		- Type of injected fluid
	Structural style		- Concentration and/or PV inj.
	Sedimentation environment		- Type of process - Miscibility
Special petrophysics	Geological age	Recovery efficiency	- Associated with Carbon Capture & Storage
	Type of seal		Number of producing wells
	°API		Number of injector wells
	Viscosity (cP)		Injection pattern:
Special petrophysics	Initial and final oil saturation (%)	Recovery efficiency	- Line, 4, 5, 7, or 9 spot
	Final oil saturation (%)		- Inverted, linedrive, or normal
	Initial pressure (bars)		- Spacing between wells
	MMP (bars)		Well density (wells/acre)
Special petrophysics	Acid number (mg KOH)	Recovery efficiency	N_{prod}/N_{inj} Ratio
	Dykstra Parsons (0–1)		Displacement efficiency
	TDS (ppm)		Performance evaluation
	Soi (%) and Swi (%)		Increased production (bpd)
Special petrophysics	Type and content of clay	Recovery efficiency	Recovery factor 1° + 2°
	Wettability		IOR/EOR recovery factor

Table 3
SOM algorithm training hyperparameters.

Initial algorithm training adjustments			
Thick		Thin	
Initial Radius	20	Initial Radius	17
Final Radius	9	Final Radius	1
Training Extension	20	Training Extension	400

between resolution and statistical accuracy (Kohonen, 2013).

There are several methods for viewing the topological space created in the SOM environment. Among them, the most commonly used methods are the unified distance matrix (U-Matrix) and the component map and the hit histogram map. The unified distance matrix is given by

$$U - Matrix = \begin{pmatrix} u_{11} & u_{11|12} & u_{12} & u_{12|13} & \dots & u_{1y} \\ u_{11|21} & & u_{11|22} & & \dots & u_{1y|2y} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{x1} & u_{x1|x2} & u_{x2} & u_{x2|x3} & & u_{xy} \end{pmatrix}$$

and gathers the average Euclidean distances of the weight vector of a neuron to the weight vectors of its neighboring neurons. The values of the elements $u_{ij|i(j+1)}$ are the distances between the neighboring neurons in the same row and column in the hexagonal grid of the map and the values of the elements to the means of the $u_{ij|i(j+1)}u_{ij}$ neighboring elements of the U-Matrix. In the color palette used in the code of the results of this work, colors closer to blue represent a greater similarity and the colors closer to red represent a greater dissimilarity. Thus, areas with low values in the U-Matrix symbolize areas with low dissimilarity or nearby neurons in n-dimensional space. On the other hand, the higher values of the matrix indicate a greater distance, that is, borders that separate neurons from different characteristics, with a high dissimilarity. The distances $u_{ij|i(j)}$ present in the diagonals of the matrix are null and represent the BMU vectors themselves in the U-Matrix plot.

The component plots show the projected distribution of the different variables in the topological space of the self-organized map. Thus, each *plot component* would be associated with the properties represented by an original variable. The topological positioning of BMUs is the same for *all component* plots generated from the projection of the trained network. This fact makes the same samples/reservoirs located in the same positions in the different *component plots*, which makes it very easy to visualize and research patterns of interest between the different variables. The application of SOM in the context of this work aims to group reservoirs and EOR projects with similar characteristics, and visualize not only their influence on other variables, with each variable represented by a component plot, but also the interrelationships of neighborhoods between the clusters of BMUs. formed, making it possible to carry out in-depth inferences and analyses.

Previous work on EOR project analysis

Research into a broad set of EOR applications can bring answers to questions relevant to the development of mature fields. Over the past decade, numerous scientific papers have developed analyses of the implementation of EOR projects on a global scale. Some of these studies addressed methodologies that underpinned the choice of EOR methods compatible with a specific type of reservoir. Such previous works, based on previous EOR experiences, bring details and lessons relevant to the industry regarding the implementation of EOR. However, the analysis and interpretation of such data in an integrated manner remain a significant challenge. The visualization and extraction of relevant and significant information from databases with a large number of dimensions are sometimes notoriously complex and complicated (Bertini et al., 2011). Moreover, the scope of the data used by these studies as a basis is sometimes limited, either concerning the EOR methods and the regions of the world that contemplate or even about the variables listed for analysis or the diversity of sources used. There is also a gap in the scientific and technological literature regarding multidimensional visualization methods of data and the integration of different forms and data sources in the field of advanced oil recovery.

3. Methodology

The sequence developed in this work was subdivided into four main stages, i.e., database structuring, pre-processing, processing, and interpretations and evaluations, based on the literature. A synthesis of the

Iterative data science CRISP-DM methodology for EOR projects data analysis

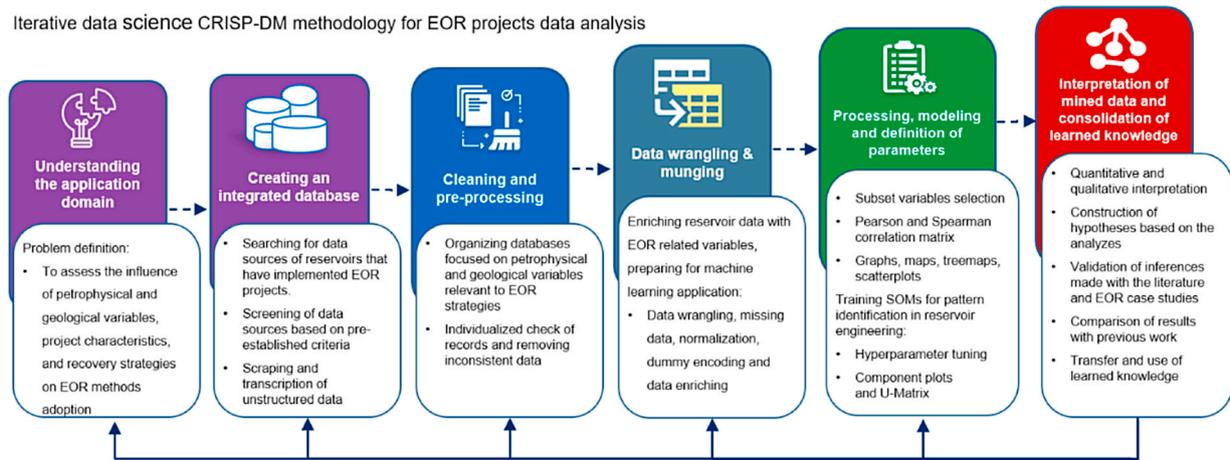


Fig. 3. Synthesis of the iterative methodological routine of data mining, which is based on the KDD and CRISP-DM models, used in the present work.

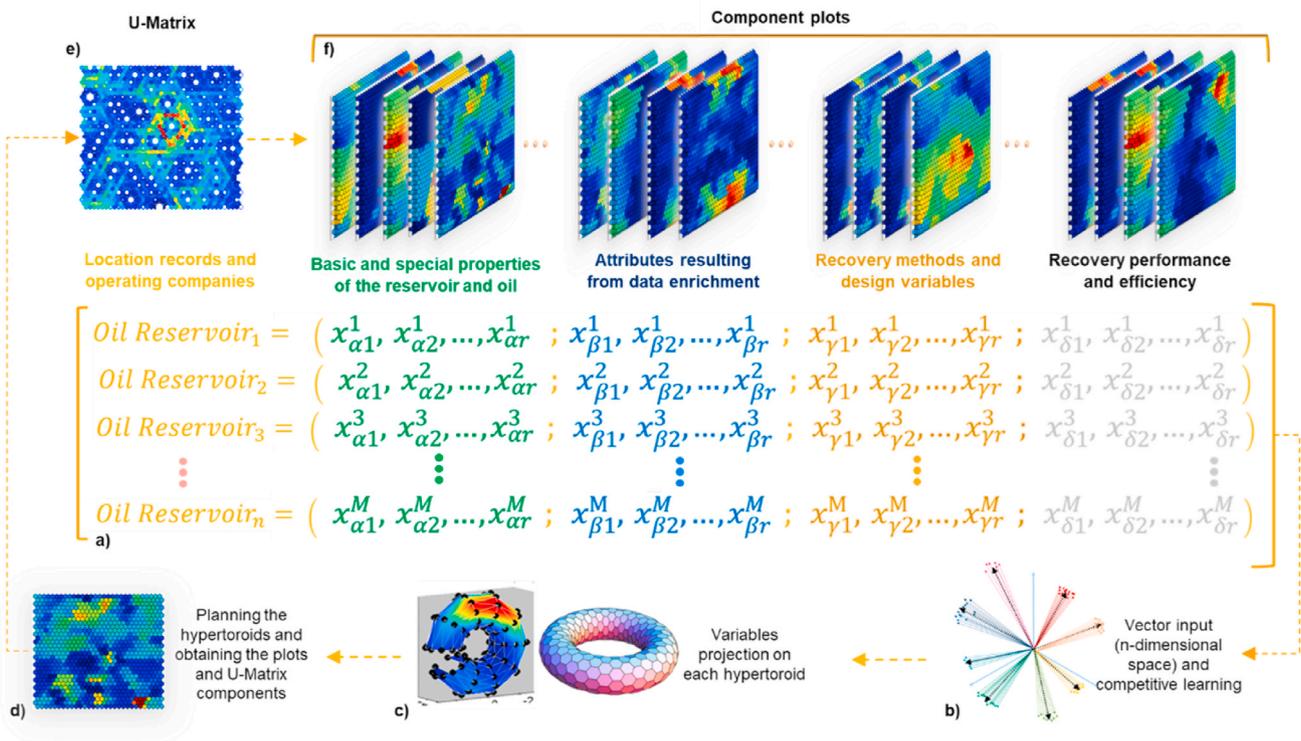


Fig. 4. Synthesis of the application of the SOM algorithm in the structured reservoirs database obtained in the present work, consisting of the steps: (a) processed database with the variables selected and modeled is introduced in the SOM algorithm; (b) the SOM algorithm is initialized, with the process of learning and adjusting the BMUs in the n-dimensional space; (c and d) the neural network has its weights projected on the hypertoroid; and the (e) U-Matrix and (f) the component plots are obtained from the planning of the respective hypertoroids. The U-Matrix represents the dissimilarities in the SOM environment of the BMUs and the component plots represent the influence of each variable for each set of reservoirs samples represented by a BMU in the SOM space.

methodological routine developed is presented in Fig. 3. This process sought to follow an iterative flow with practical improvements suggested by methodologies such as Knowledge Discovery in Databases (KDD) and Cross Industry Standard Process for Data Mining (CRISP-DM), developed by IBM to ensure that the data used is relevant and adequately processed to address the proposed issues (Gregory Piatetsky, 2014; Huber et al., 2019; Shafique and Qaiser, 2014; Wirth and Hipp, 2000).

Source and capture information to make up the database

As summarized in Table 1, the main data sources used can be

subdivided into.

- i. Scientific articles considered references in the application of EOR. Certain articles explain the state of the art in the application of certain EOR classes; others, in turn, focus on reporting experiments from production fields where EOR methods have been implemented.
- ii. Comprehensive surveys and studies of the application of EOR methods in reservoirs of certain specific regions of the world or certain sedimentary basins, such as the North Sea (Awan, 2008) of shale oil Chico Sambo, carbonate, offshore reservoirs and fields in Brazil, like Peregrino in the Campos among others. (Alvarado

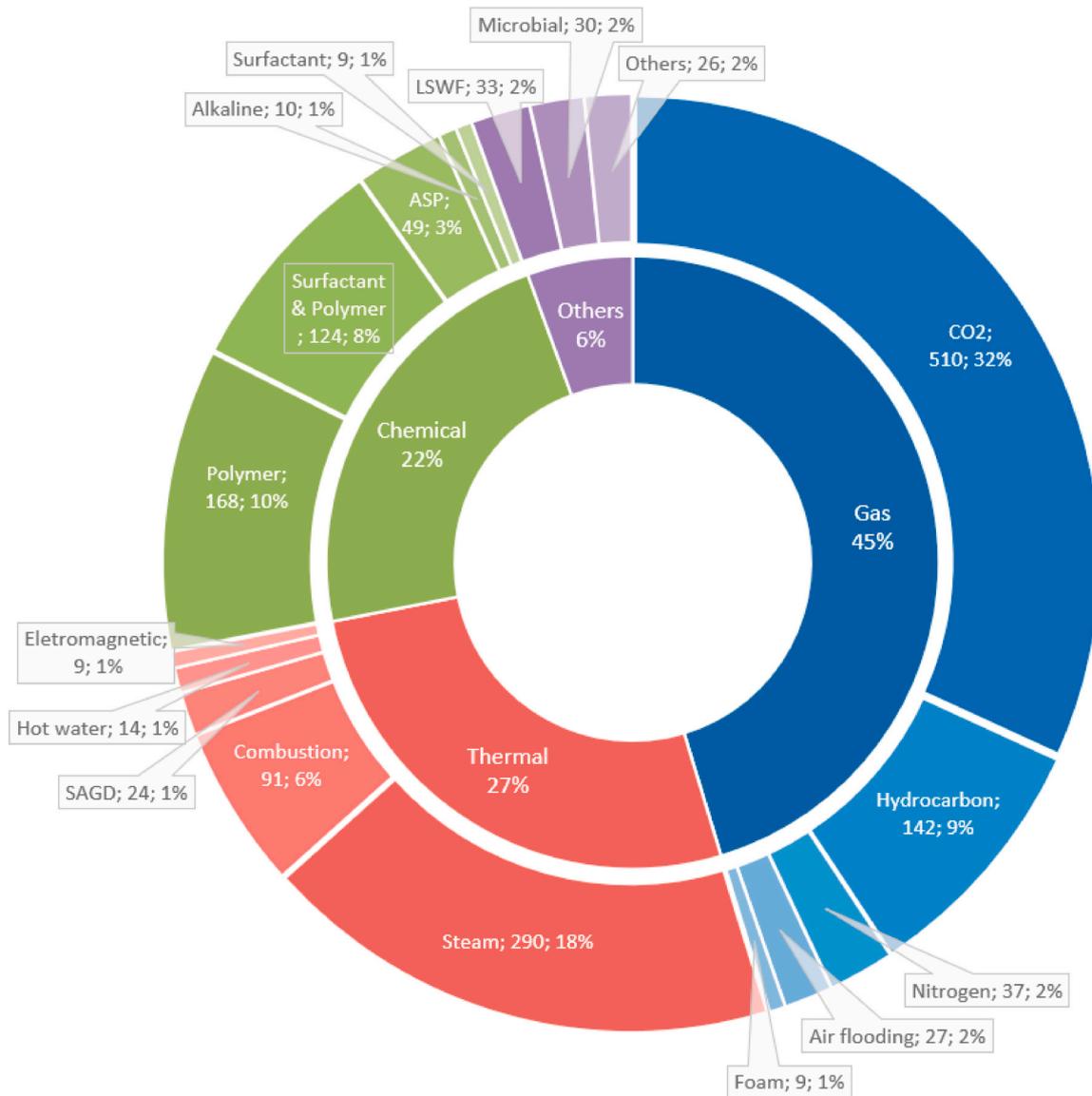


Fig. 5. Classes of EOR methods that are present in the database.

- and Manrique, 2010a, 2010b; Bondor, 2005; Manrique et al., 2008; Raney et al., 2012).
- iii. Set of field experiments on a specific method, such as CO₂ injection associated with carbon capture and storage (Karacan et al., 2023; Lake et al., 2019; Warwick et al., 2019), alternating injection of water with gas (Awan, 2008; Gbadamosi et al., 2018b), use of microbial EOR methods, low salinity water injection (Chavan et al., 2019; Chen et al., 2021; Le et al., 2015; Patel et al., 2015; Wang and Fu, 2018), injection of chemicals (AlSofi et al., 2019; Gbadamosi et al., 2019b; Sheng, 2010), in-situ combustion (Miller, 1995; Turta et al., 2007), among others.
 - iv. Articles with broad information, but less structured or with specific data (J. J. Brodie et al., 2012; Da Costa Fraga et al., 2015; De Moraes Cruz et al., 2016; Jin, 2017; Lamas et al., 2018; Morel et al., 2015), among others;
 - v. Reports and databases of international agencies and institutes, such as the International Energy Agency (IEA), Center for Strategic and International Studies (CSIS), Global Carbon Capture and Storage Institute (CCSI), OPEC (Organization of the Petroleum Exporting Countries), among others;
 - vi. Regulatory and research bodies from several countries;
 - vii. Presentations to investors and technical reports of operating companies in the fields; magazines and news sites specialized in the oil and business sector;
 - viii. Books containing case studies and real examples of the application of EOR (Alvarado and Manrique, 2010b; Sheng, 2010; Vishnyakov et al., 2020).
- The use of various data sources in this work occurred mainly for the reasons listed below: (i) in the data-driven approach, there is the insertion of information from different sources and projects from different localities, in which the diversification of sources mitigates possible bias; (ii) the presence of a large number of variables causes a large sparsity of data in multidimensional space, which may lead to the effect known as the curse of dimensionality and in this case, obtaining more data reduces the percentage of missing data; (iii) data sources in isolation do not have all the information about the respective EOR projects they address but tend to be complementary; (iv) by using different sources, it is possible to expand the representativeness and scope of the database given the total universe of EOR projects existing in the world (Aggarwal, 2015; Bertini et al., 2011; Ganguli and

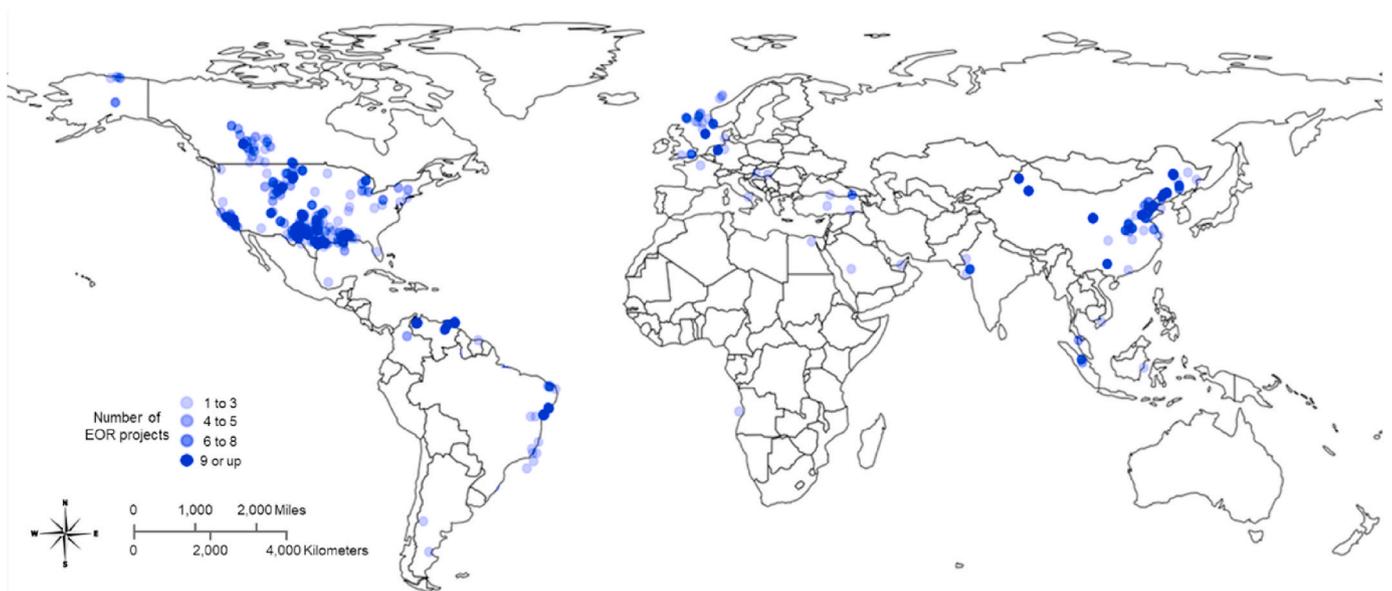


Fig. 6. Location of the oil fields that had EOR applications present in the database presented here.

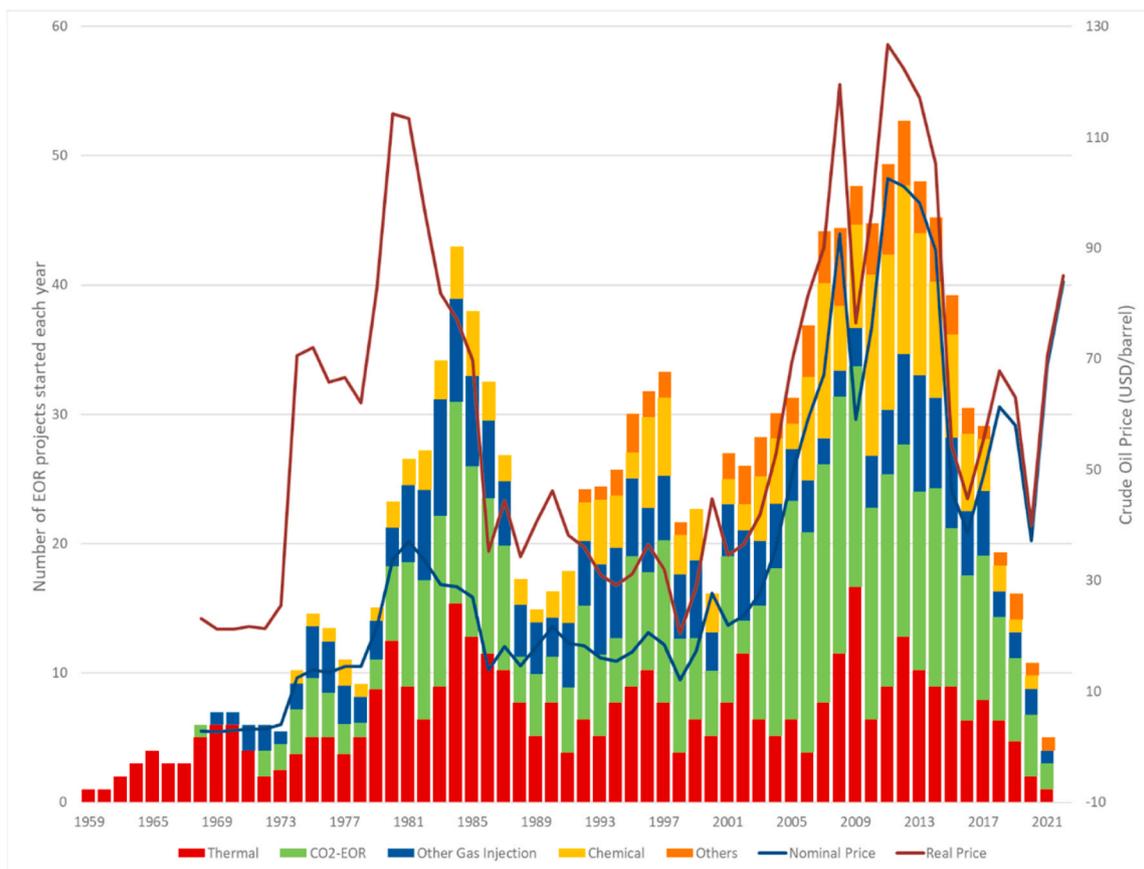


Fig. 7. Project frequency bars for each EOR class over the years overlapped by the curve relative to the price of Brent oil barrel, nominal and debased on inflation, based on the US Economic Index.

Sompolinsky, 2012).

Structuring of the database

The compiled and processed Structured Query Language (SQL) database contains 1237 instances, each referring to an EOR project

around the world, with 323 variables that can be analyzed. In structuring the database, each instance of the data set represented the execution of an EOR project in an oil reservoir in the world, having as variables four major sets of characteristics: (i) location records, (ii) lithological, (iii) petrophysical, (iv) project characteristics, constituting real data of the implementation of EOR in the world, bringing with it

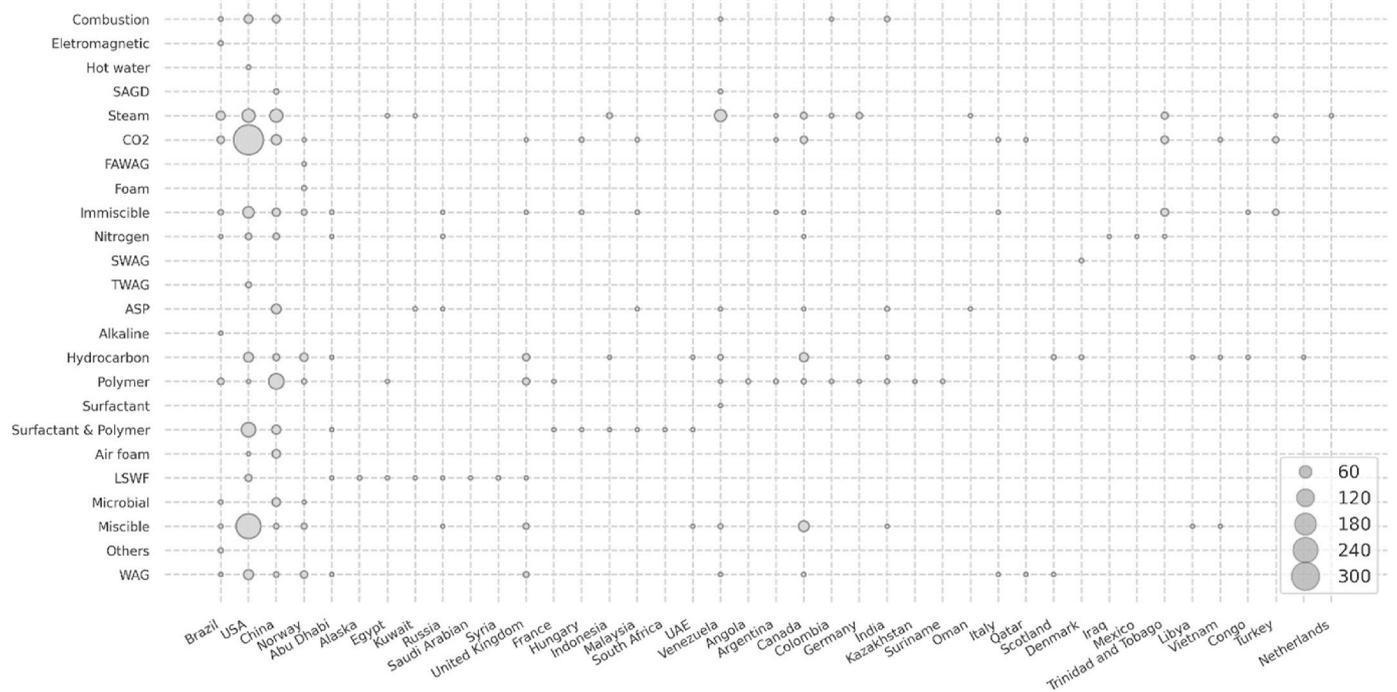


Fig. 8. Categorical scatterplot with the distribution of subclasses of EOR methods vs. the countries in which they were applied. The size of the points is proportional to the accumulated frequency of projects of the method to the respective country.

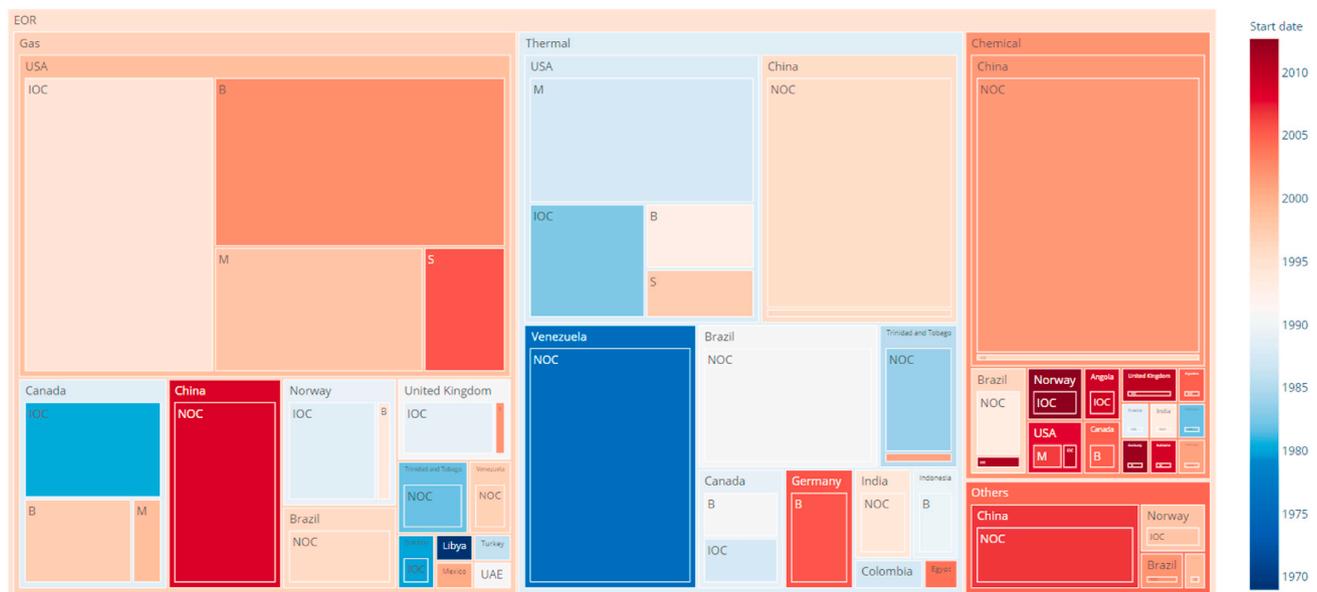


Fig. 9. Treemap represents the distribution of EOR projects, grouped according to EOR class, the country of location, and classification of the field operating company, according to IOC (International Oil Company), Big (annual revenue between 5 and 20 billion dollars), Medium (annual revenue between 500 million and 5 billion dollars), and Small (annual revenue below 500 million dollars). The color of each rectangle refers to the average start date of the projects, indicated by the scale on the side.

relevant information about the challenges, technical difficulties, and other influencing factors.

Preprocessing

3.1.1. Cleaning & variable management

To make the database as consistent as possible, the data were pre-processed (Table 1), involving a sequence of steps. Initially, noise removal was made, including spurious and/or invalid information and

the standardization of variables, and units were converted to gather them into a common numerical identity (from feet to meters, km² for acres, etc.). Data from duplicate reservoirs were merged or removed. In a general inspection of the data set, there were possible inconsistencies or discrepancies and outliers were verified. Finally, data were removed that indicated inconsistencies, with checks developed from more detailed and individualized research on the respective fields.

The variables obtained and the overall structure of the processed database are shown in Table 1.

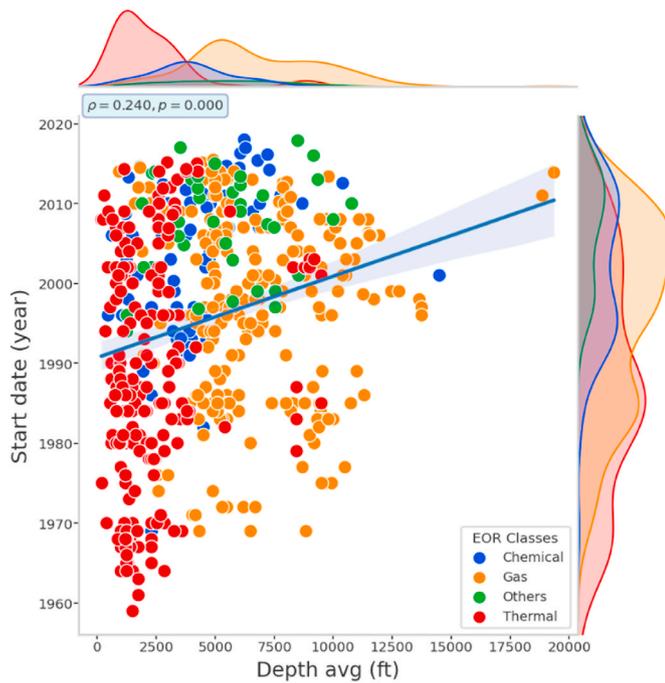


Fig. 10. Scatter plot of the variable depths of the reservoirs and start dates of the projects, with the categorization of the data as to the class of the EOR method used.

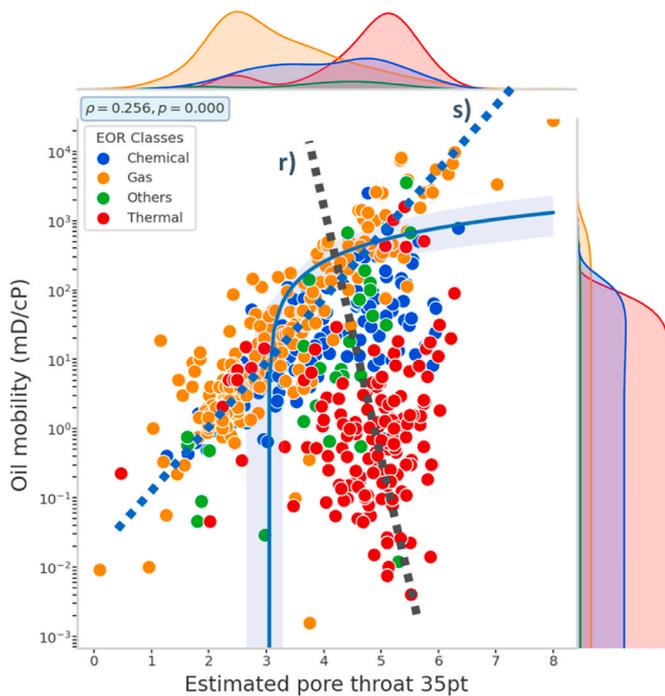


Fig. 11. Cross plotting of the variables estimated oil mobility and estimated pore throat size. The lines r) and s) indicate distinct correlation trends for data subgroups.

To make the structure of variables appropriate to the application of AI techniques, the variables were managed through operations like the normalization of variables to avoid differences in the training weights of the RNAs, the transformation of categorical variables using the one hot encoding technique. The creation of new variables with physical attributes or quantities of interest from the combination of previous properties was also carried out, favoring new parameters and enriching the

analysis. Examples of this data enriching process are the variables estimated pore size and estimated oil mobility, discussed in section 4.1.2. The application of logarithmic transformation for variables that presented a log-normal distribution for variables such as permeability, viscosity and oil mobility, density and number of injector and producer wells.

Processing and analysis methodology

3.1.2. Multidimensional analysis by self-organizing maps

The integrated database with all acquired reservoir data, after undergoing cleaning and pre-processing, was fed into the SOM algorithm, where an n-dimensional random data space was created. A hexagonal grid was selected for the visualization format, and a toroidal surface was utilized for projecting neurons or BMUs to maintain similarity and topological relationships while avoiding edge effects. A map size of 20 x 20 cells was chosen to allow for adequate segmentation and identification of patterns within each reservoir group formed by the algorithm and visualization in the component plots. The initialization was set to random, and the initial weight adjustments during training were according to Table 2. Fig. 4 illustrates the systematic approach applied in the SOM analyses.

After the generation of the self-organized map, the representative 2D images of the n-dimensional space given by the component plots and the U-Matrix are produced. The component plots make it possible to visualize and quantify the contribution of the reservoir oil and petrophysical properties and production attributes variables analyzed for each resulting neuron in the self-organizing map, making it possible to verify interrelations between the responses of the analyzed variables. The U-Matrix then allows the classification of the data related to the similarity vector, which is constructed from the original instances.

Each neuron that has associated reservoir instances on the self-organized map was represented by a BMU, that is, a "mini-cluster" of the input instances, grouping reservoirs or production fields with similar characteristics. As a result, we will have the BMUs for each reservoir analyzed.

Interpretations and validations of observed patterns

As part of the analysis of the results generated, an integrated procedure was performed, making use of both quantitative and qualitative interpretations. Scatterplots with reservoir data distribution, histograms, and correlations Pearson, Spearman and statistical significance between variables are used jointly along with the observation of component plots generated by SOMs. Such component plots show the structure of groups of reservoirs formed by the U-matrix, with their representation in the different component plots. Based on possible observed patterns, hypotheses are constructed based on the joint analysis of the variables, and then a search is made in previous literature on reservoir engineering and application of EOR in point fields that validate or refute such observations identified in the results generated. This process is carried out cyclically, as recommended by the CRISP-DM and KDD methodologies, with a continuous expansion of the database, reprocessing, modeling, and data analysis.

4. Results

Overview of the database

The resulting integrated database was able to integrate 323 variables relating to the 1237 reservoirs operated by 96 companies distributed in 36 countries. As shown in Fig. 5, the collected projects include the application of 4 large classes and 24 EOR subclasses.

4.1.1. Location and history of EOR projects

The existence of EOR application projects with significant global

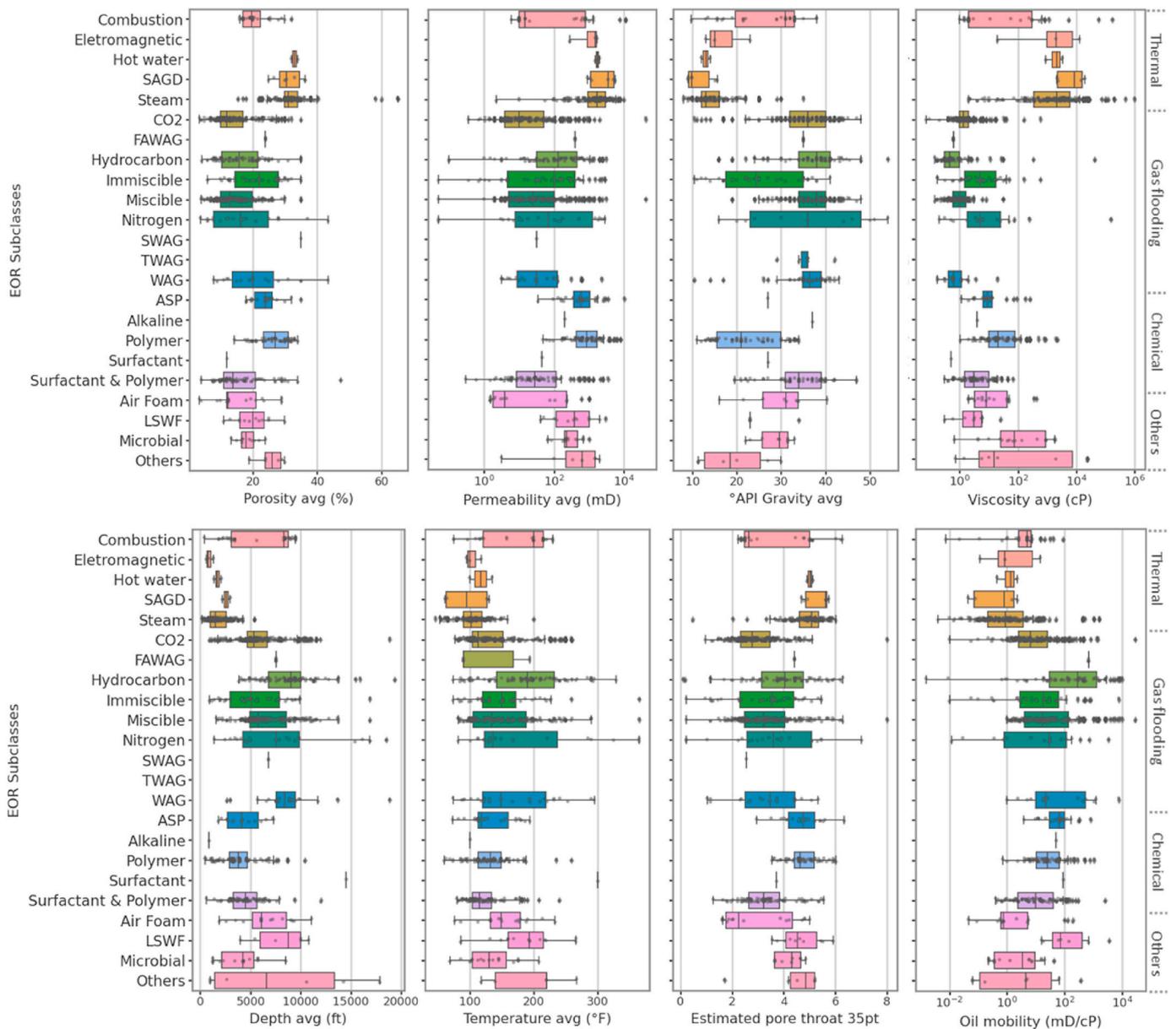


Fig. 12. Boxplots for the main types of EOR methods for different variables of a technical and petrophysical nature.

distribution was observed, revealing the dissemination of EOR techniques. However, as illustrated by Figs. 6 and 8, a concentration of the applications of these techniques in regions involving certain countries is noted. Among these regions, we highlight the USA, Canada, China, countries bathed by the North Sea, Venezuela, and Brazil, among others. The most frequent regions are distributed from North America and Europe, with high energy demand and with an already large number of EOR projects implemented decades ago, passing through China, with an intense recent application, to countries in Latin America, the Middle East, and Africa with a relatively smaller scale of adoption of advanced oil recovery techniques.

Fig. 7 brings the frequency of projects for each class of the EOR method over the years between 1959 and 2015. The distribution overlapped the time series of nominal oil prices and debased inflation based on the US. The discounted price of inflation was also included to obtain a fairer comparison between different economic scenarios. It is noted that the frequency of the start of EOR projects generally accompanies the price curve of petroleum. However, between the 1980s and 1990s, the frequency remained high despite the fall in the price of oil. This

variation was probably due to the U.S. Department of Energy incentive policies for the implementation of EOR projects.

In addition, a low frequency of projects with an initial date above 2014 was noticed. However, it is not possible to affirm that in the last decade, there have been few EOR projects implemented. The low frequency found may simply reflect that the most recent data on the application of EOR have not yet been disclosed by the operating companies or published in scientific articles. Thus, these more recent projects were not considered in the structuring of the database at present. Because it was not possible to obtain the final dates of the implementation of the EOR projects for a considerable part of the fields, we opted for the frequency of EOR projects per year in the chart to take into account only their respective start dates.

Some classes of EOR methods have greater application in certain countries, as can be seen in Fig. 7. In particular, it is noted that the U.S. leads the application of steam injection, followed by Venezuela, China, Canada, and other countries. The largest number of projects and the highest oil production stemming from the implementation of EOR methods occurred in the USA. Among the reasons for this, we highlight

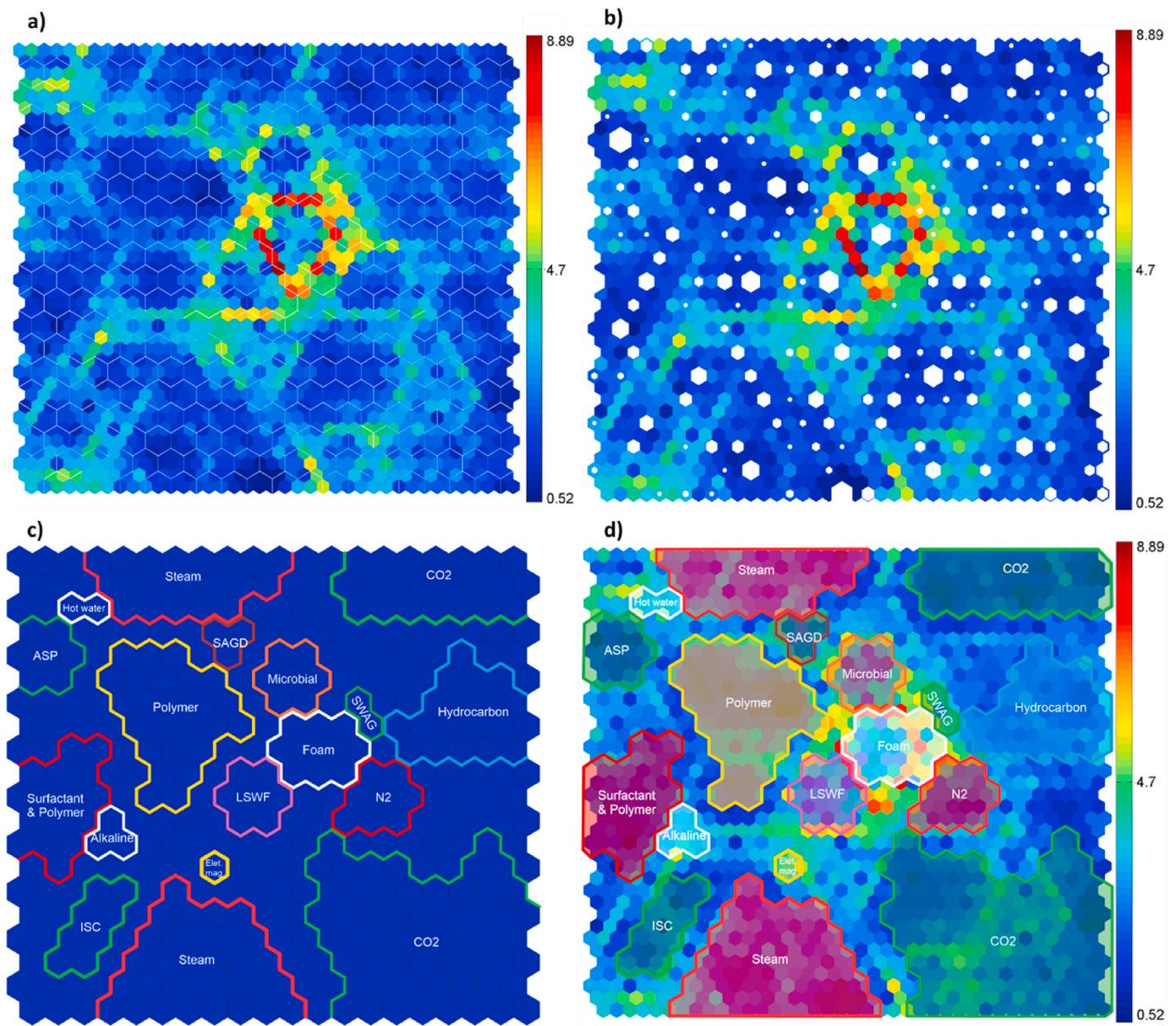


Fig. 13. U-matrix resulting from the training carried out with the data of EOR projects with four levels of highlights: a) U-matrix with the representation of the hexagonal network of neurons, where each external white contour of the hexagon represents a neuron of the artificial neural network, and in which the hexagons adjacent to the edge indicate the measure of dissimilarity. The warmer colors indicate high dissimilarities of the neuron in question with its adjacent ones, and the colder colors have greater similarities; b) U-matrix with hits represented in white color. The size of the hexagon is proportional to the number of reservoirs that the respective BMU is representing; c) interpreted lines related to the regions of neurons most influenced by the data of the respective EOR method, and d) overlap of the regions of the lines interpreted in c) in the U-matrix image of a).

the maturity of the U.S. in the application of EOR methods, with a large number of depleted oil fields and state and federal incentive policies. In addition, the country has a high availability of technology and infrastructure, making it the ideal place for the implementation of pilot and commercial EOR projects, even in periods of low oil barrel prices.

China, on the other hand, has several projects based on chemical methods. Observing together the graphs in Figs. 3 and 7, in particular, there is an increase in the number of publications and EOR projects related to the injection of chemicals, in particular, the polymer injection and surfactant. While in the decades of 1980 and 1990, there was a maximum of 5 chemical EOR projects per year, in the 2000s and 2010s this number came to overcome 15 projects. Towards the frequency of publications along the years, such growth becomes even more evident. Such a phenomenon can be explained, especially in China, by regulatory incentives, lower costs of chemical products, and the development of

new compounds that motivated the study and application in various reservoir conditions. This fact reveals both the use of chemical methods on a larger scale, as well as Chinese leadership in such a EOR class, in recent years (2000). For gas injection, US leads the number of projects executed, followed by the North Sea countries.

The EOR projects collected on a world scale were or are carried out by several operating companies in the oil industry. They are globally known, and EOR projects collected worldwide were or are conducted by several integrated companies throughout the oil production chain and derivatives trading, such as British Petroleum (BP), Chevron, China National Offshore Oil Corporation (CNPC), Equinor, ExxonMobil, Petrobras, Petronas, etc. There are also independent and often regionally active companies such as Anadarko, Apache, Athabasca, Canadian Natural Resources (CNRL), Carrizo, Continental Resources, etc., which represent a significant share of the leadership in the execution of EOR

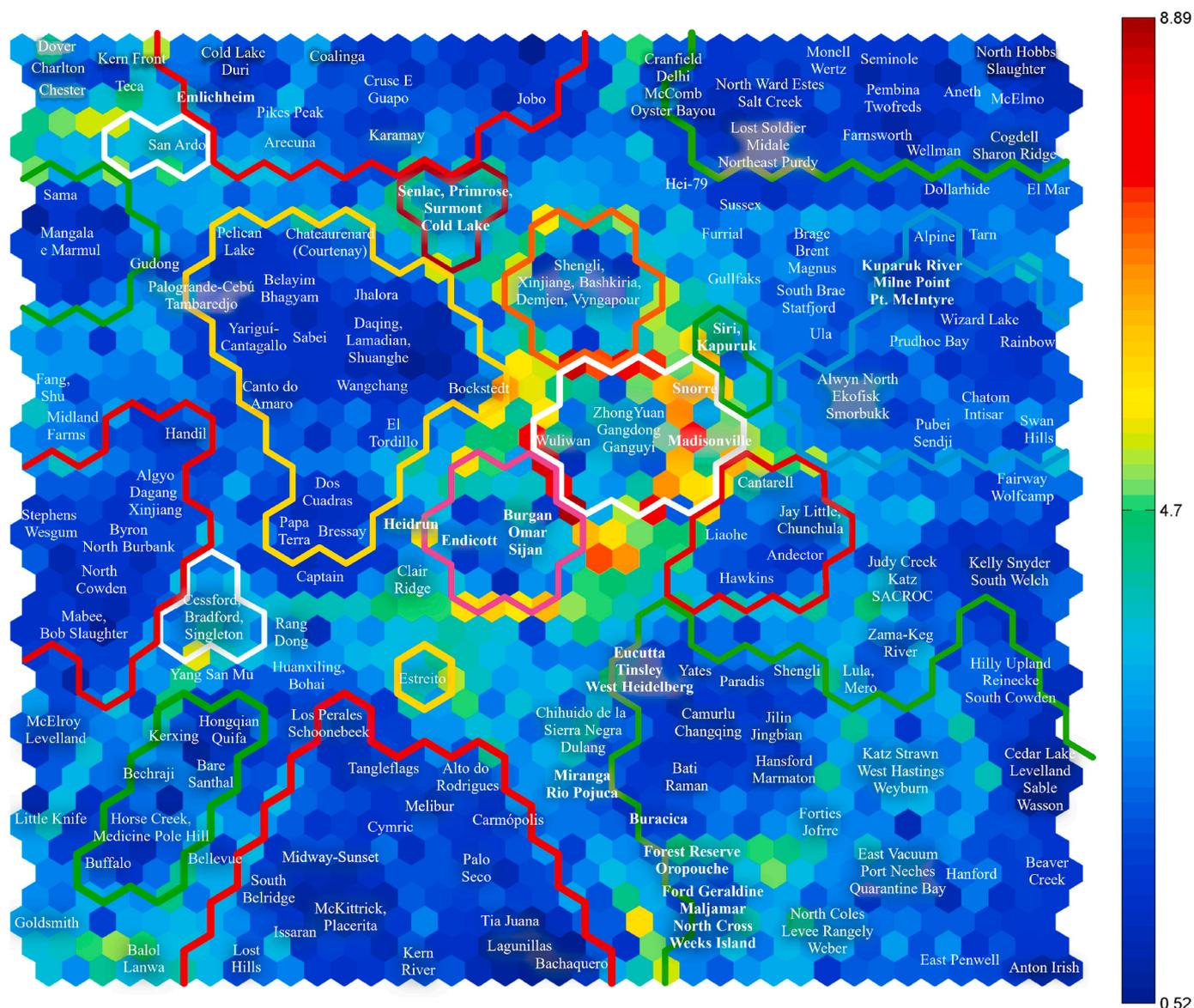


Fig. 14. U-matrix resulting from the training carried out with the data of EOR projects. Examples of projects representative of the respective BMUs have their names presented.

projects.

From the treemap in Fig. 9, it is possible to observe the tendency of applications of chemical methods being carried out mainly by integrated companies, whether National Oil Companies (NOCs) or International Oil Companies (IOCs). On the other hand, projects involving gas injection methods have been executed not only by integrated companies (NOCs and IOCs) but also by a significant proportion of companies classified as medium and independent in the Exploration and Production (E&P) sector. For the thermal methods, there is a diversity of companies operating, like NOCs and IOCs, and a lower proportion of small and medium-sized independent operators focused on E&P.

4.1.2. Reservoir parameters

Based on the scatterplot in Fig. 10, different behaviors are noted for the EOR classes regarding the relationship between the average reservoir depth and project start dates. First, a positive general correlation between more recent projects with higher reservoir depths can be noted. Secondly, such a positive correlation is observed for chemical and gas injection methods. On the other hand, the projects of thermal methods present a distinct behavior, showing an apparent barrier in the advance

to higher depths when compared to the other classes of EOR.

From petrophysical parameters, the synthetic parameter logarithmic ratio between permeability and porosity was calculated, which is related to the estimated size of the pore throat and its lithology (Jerauld and Salter, 1990; Kolodzie, 1980; Rezaee et al., 2006). Based on Fig. 11, it is noted that thermal methods tend to be applied in reservoirs with the estimated throat of higher pores and oils of lower and more viscous mobility indexes when compared to chemical and gas injection methods. Gas and chemical injection methods have greater application in lighter oils with smaller pore throats and chemical methods at intermediate values for these variables (Jerauld and Salter, 1990; Kolodzie, 1980; Rezaee et al., 2006).

Two trends of data disposition, represented by the lines r and s, are noted. There is a tendency for gas-instilled reservoirs to be available along the dashed line in purple, with reservoirs that have had chemical methods just below this line; thermal methods tend to be available along the black r-line. This fact indicates that for gas injection, a larger estimated size of the pore throat tends to result in a greater oil mobility of these reservoirs. For thermal methods, there is an inverse relationship: a larger pore throat results in a lower oil mobility of these reservoirs. By

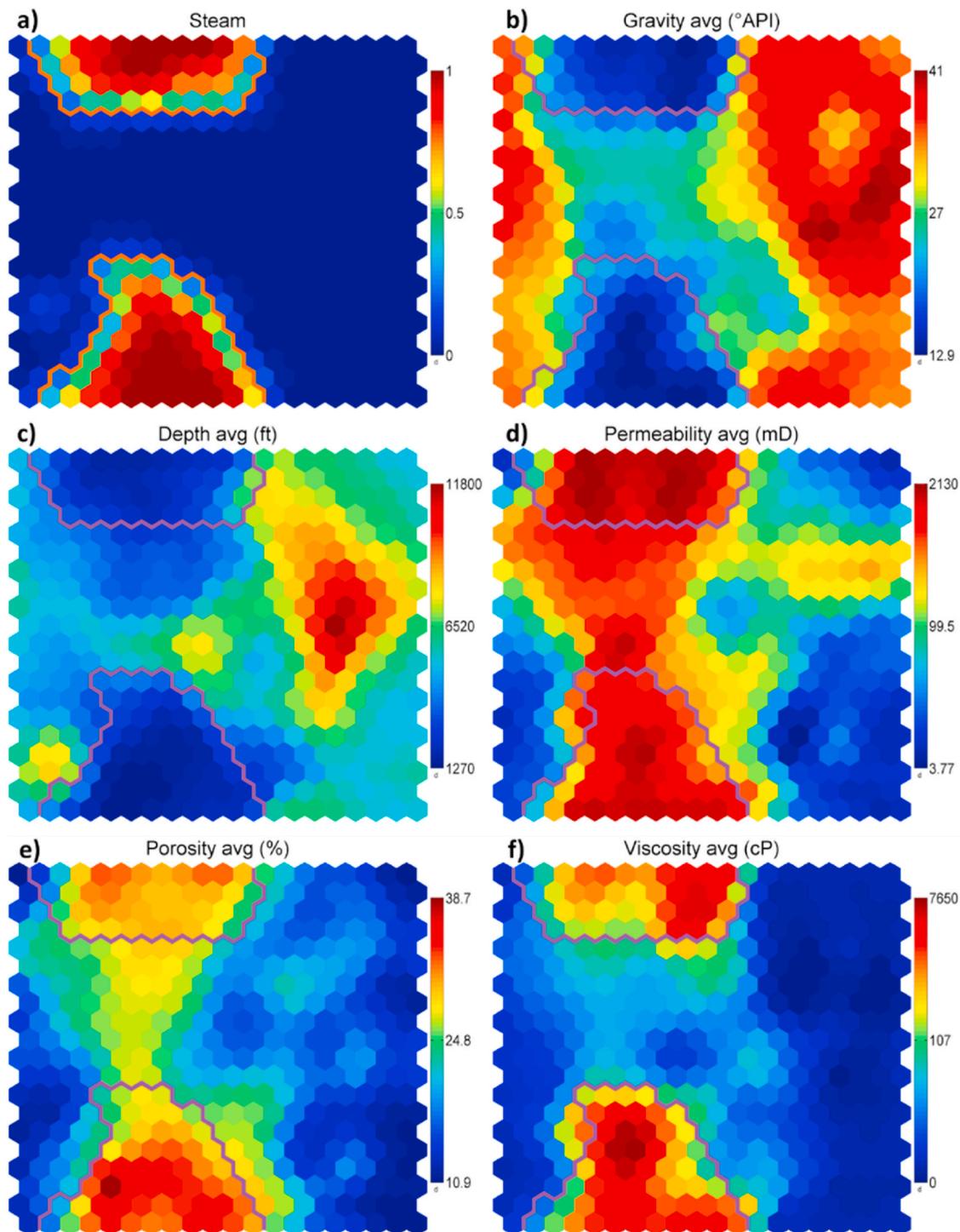


Fig. 15. a) *Component plots* referring to the steam injection method; b) to f) represent *component plots* of the other basic variables of the reservoirs. The highlighted line delimits the neurons activated by the steam injection method.

inspection, data groupings according to the dominant lithology of each reservoir are noted when correlating the porosity variables and estimated pore throat size. In particular, sandstone tends to present higher porosity values and estimated pore throat size, while carbonates have a greater dispersion of the data, but with a concentration in lower values of porosity and estimated pore throat size. In this sense, steam injection projects tend to be applied in reservoirs in general sandstone with high estimated pore throat sizes and permeabilities but which have high viscosity oil, which entails a low mobility index, as shown in Fig. 11.

The boxplots in Fig. 12 summarize the distribution of part of the

technical variables concerning the EOR methods belonging to each main class. It is possible to observe that for the same class of EOR, there are significant variations in the application ranges of the technical variables according to the EOR method used. Such differentiations between recovery methods will be addressed in the section "SOM analysis of EOR methods".

SOM analysis of EOR methods

SOM training obtained final and topographic quantization errors of

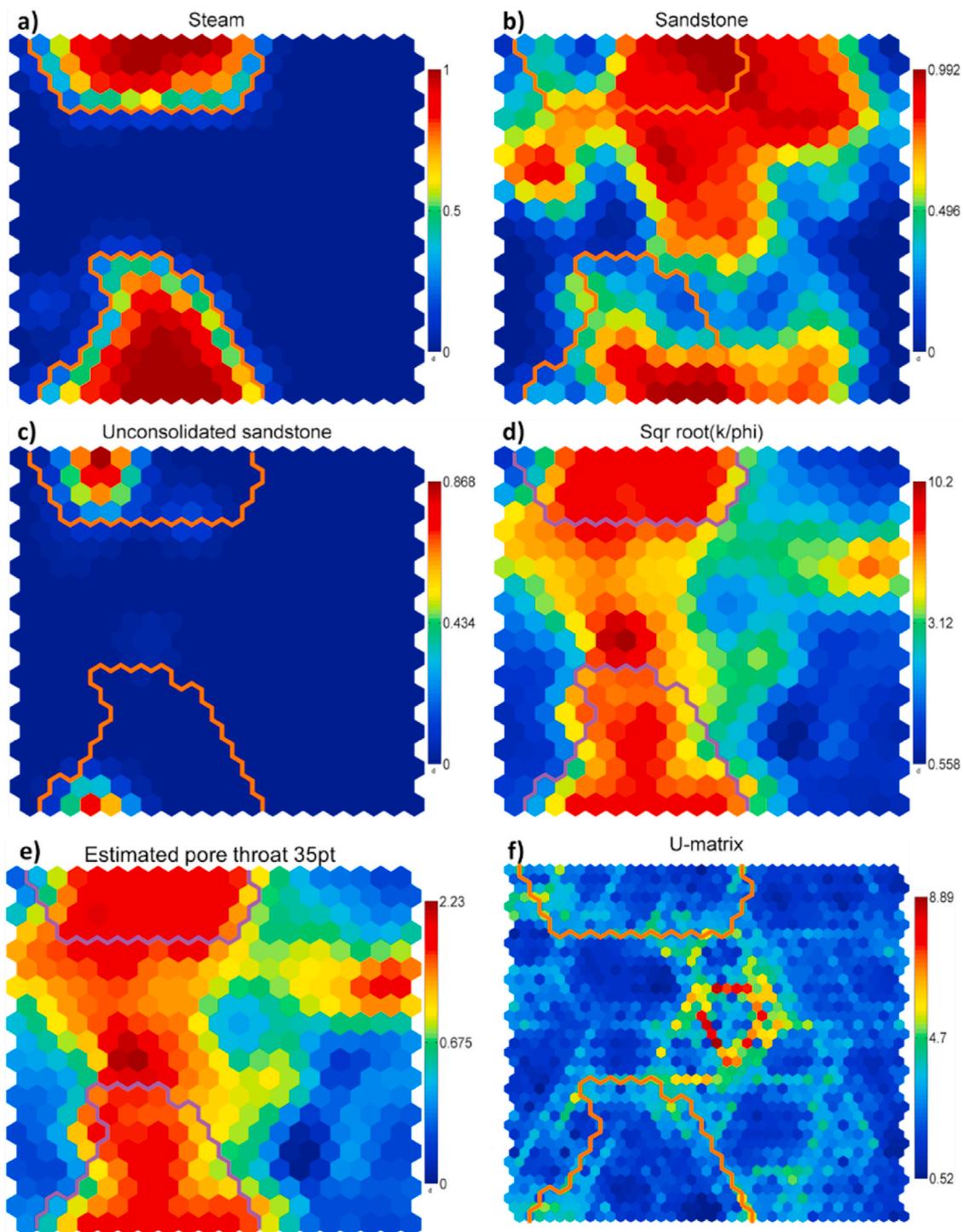


Fig. 16. a) *Component plot* referring to the vapor injection method; b) to f) represent *component plots* of variables related to the lithologies of the reservoirs. The highlighted line delimits the neurons activated by the steam injection method.

0.193 and 0.075, respectively. The U-matrix resulting from the application of the algorithm in the data set can be observed in Fig. 13. A good dispersion of the sample data sets was observed by the U-Matrix, which suggests good representativeness of the data groups in question. Fig. 14 illustrates the U-Matrix with the names of examples of EOR projects represented by each of the BMUs.

4.1.3. Thermal methods

4.1.3.1. Continuous, cyclic steam injection and SAGD. Steam Flooding

(SF) steam injection projects account for 18% of the database, with a total of 230 projects cataloged in 16 countries. It is still the main thermal method, and most projects are mainly in the USA, Venezuela, China, Brazil, and Canada. The first cataloged projects of SF began in the late 1950s in the US, followed by Venezuela and Canada. In the following decades, it also had applications on a field scale in Brazil, Trinidad and Tobago, China, and India in reservoirs with viscous oil. In China currently, about 80% of the production of fields with heavy oil comes from the use of steam injection. The cataloged projects are diverse with different configurations of the water and steam injection processes, with

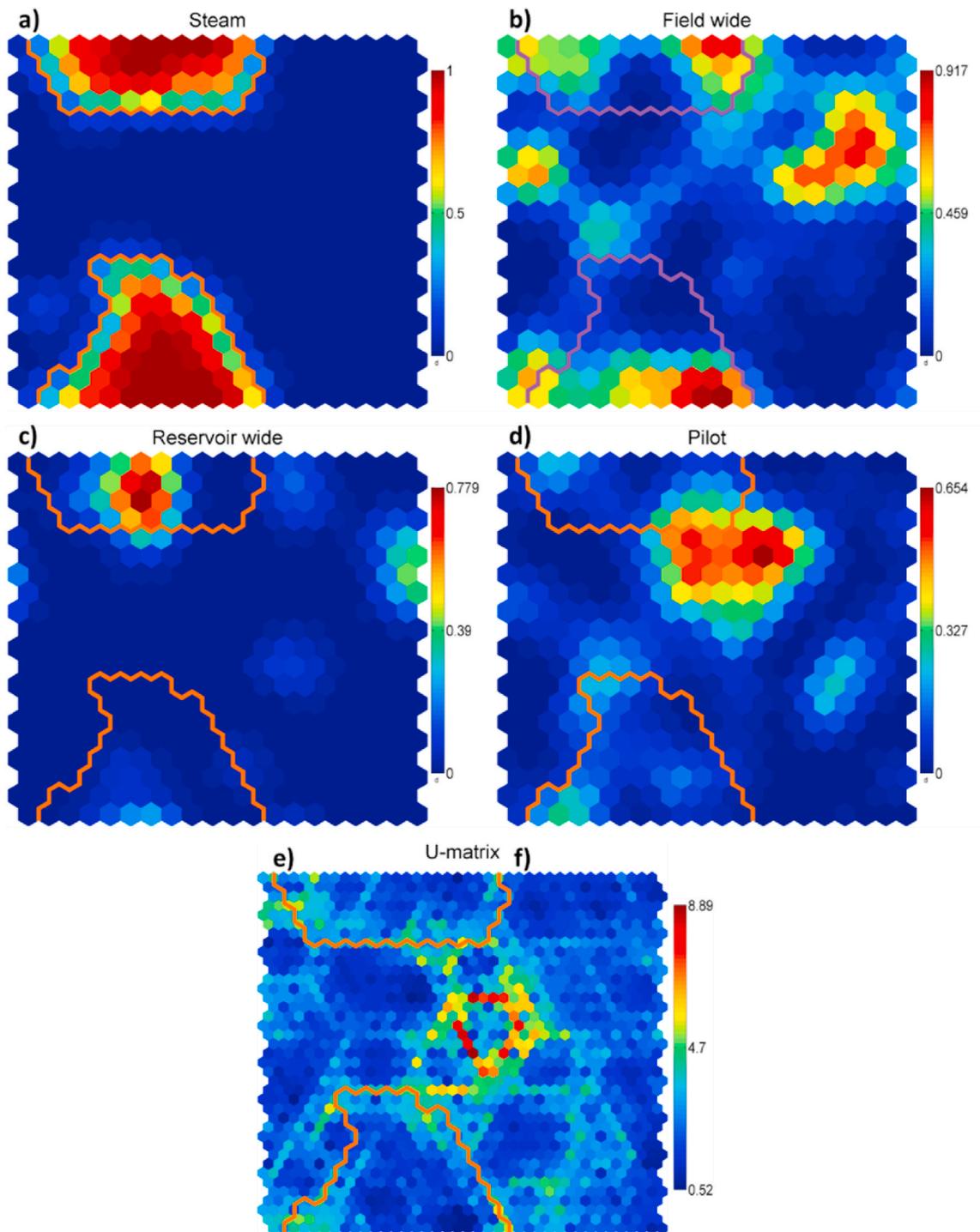


Fig. 17. a) *Component plot* referring to the steam injection method; b) to d) *component plots* referring to the project scale (whether pilot, reservoir scale, or field-scale); e) the U-matrix. The highlighted line delimits the neurons activated by the steam injection method for vapor injection.

most corresponding to (i) continuous steam injection, followed by (ii) cyclic/batch injection of steam, (iii) hot water injection, Hot Water (HW), and (iv) steam injection into directional wells, Steam Assisted Gravity Drainage (SAGD). In particular, Canada was the first country to conduct an SAGD project in the fields of Christina Lake and MacKay River.

Based on the component plots of Figs. 15 and 16, the sets of neurons with the greatest presence of projects associated with steam injection were analyzed. These neurons have mean porosity values of 27% (but with clusters presenting 16%–37% porosity, permeabilities of, on

average, 1000 mD, low depths of 3000 ft, °API mostly up to 17, viscosity above 200 cP, average estimated pore throat values above 3.3). Through inspection of the reservoirs present in such neurons, outliers with permeabilities of the order of 10 mD, such as the Diatomite formation of the South Belridge field (MURER et al., 1997), are noted. Among the steam injection projects, those that used SAGD had applications in reservoirs with average porosities of 22%–37%, permeabilities from 900 to 6000 mD, viscosities from 1000 to 14000 cP, °API from 9 to 19, and depths from 2000 to 3200 ft. Thus, projects with the use of SAGD tend to have applications in heavier oils and less thick formations due to directional

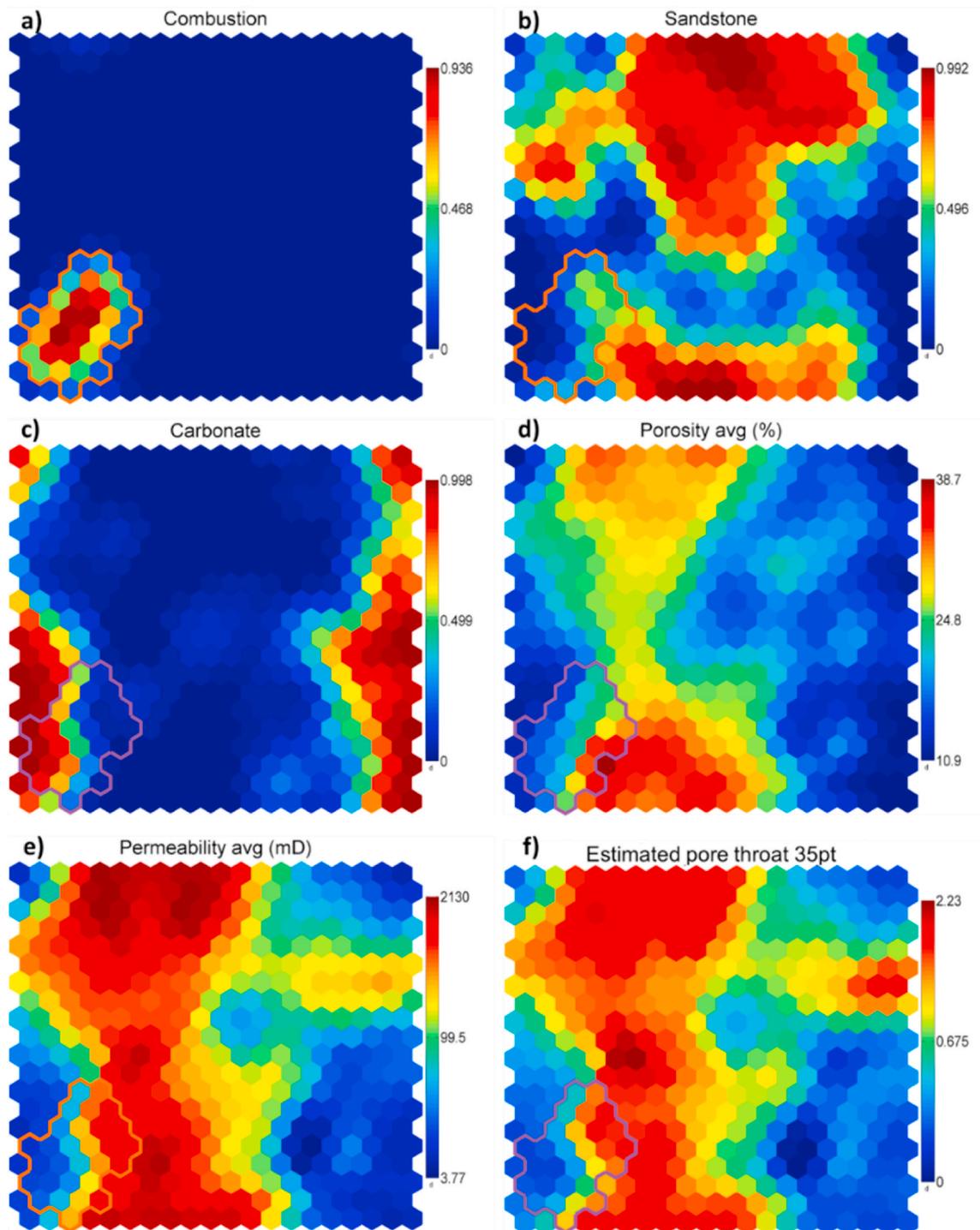


Fig. 18. a) Component plot referring to the in-situ combustion method; b) to f) represent component plots referring to the project scale (whether pilot, reservoir scale, or field-scale). The highlighted line delimits the neurons activated by reservoirs associated with the in-situ combustion method.

wells and are more favorable to oil flow when compared to the rest of the steam injection projects (Murer et al., 1997).

According to the component plots in Fig. 16, the steam injection projects had their application mostly in sandstone reservoirs. In the 9 applications cataloged in carbonates, the reservoirs have lower values of permeabilities (in the range of 10–100 mD) and porosities (in the range of up to 20%), but they are considered high values for formations of carbonate matrices. In addition, such carbonate steam injection projects have °API above 16, which is higher than for steam injection projects in sandstone. Among the sandstone reservoirs, 38 projects were used in

unconsolidated sandstone, which presented higher mean values for porosity and permeability and lower depths when compared to reservoirs with consolidated sandstone. Such trends of lower values of °API and higher values of hydraulic properties for projects in sandstone reservoirs can also be observed for the HWF and SAGD methods.

According to the component plots in Fig. 17, it is noted that most steam injection projects already have field-scale applications. In addition, it is the only thermal method in which offshore projects have been identified. There are 7 projects located, for example, in Bohai Bay, China, and Congo.

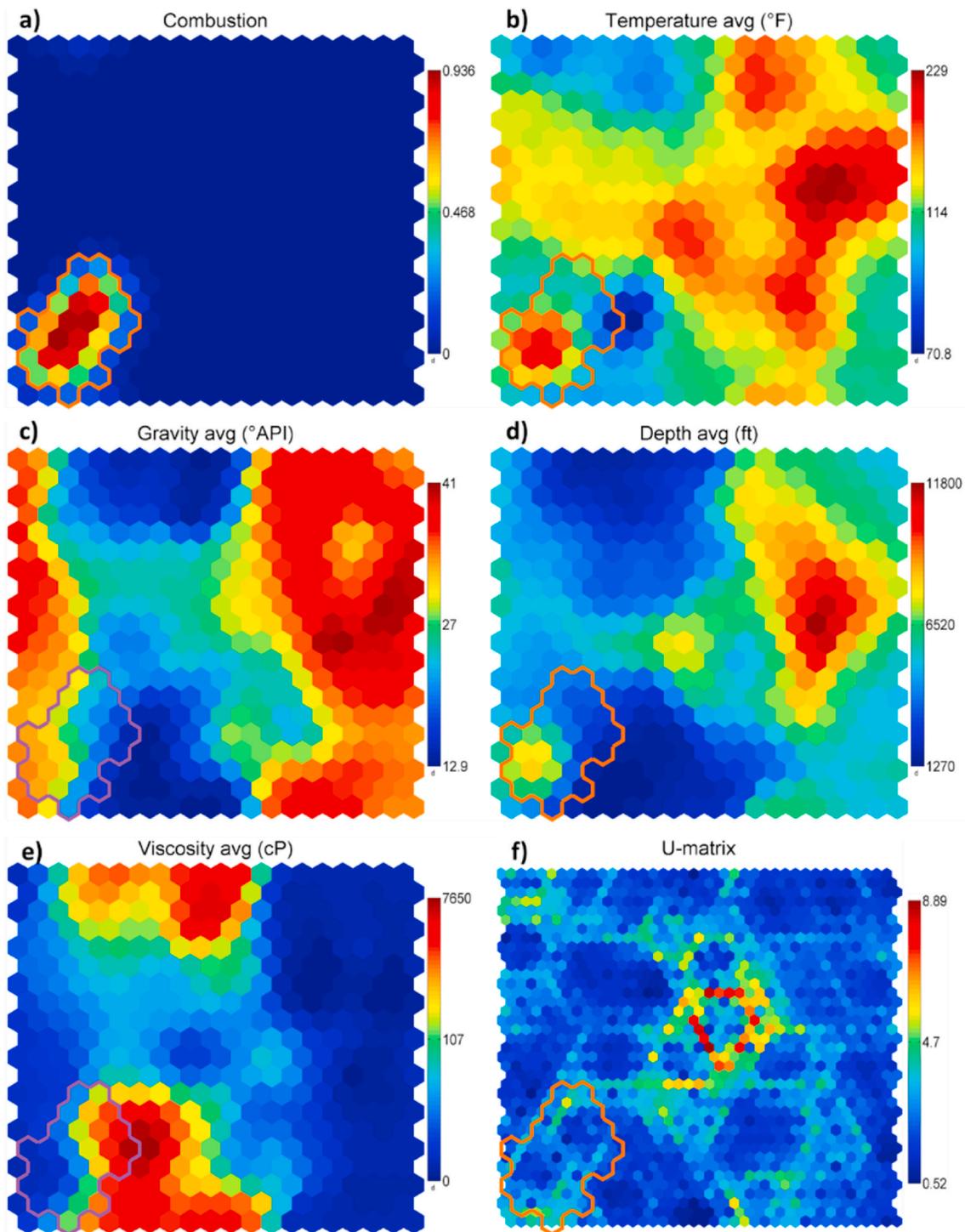


Fig. 19. a) Component plot referring to the in-situ combustion method; b) to f) represent component plots of the other basic variables of the reservoirs. The highlighted line delimits neurons activated by the in-situ combustion method.

4.1.3.2. In-situ combustion. In-situ combustion (ISC) projects represent 6% of the database, with a total of 91 projects cataloged in 7 countries. Most projects are mainly in the USA and China. The first cataloged projects began in the late 1960s in the U.S. and Brazil, with wide application in the following decades also for India, China, Venezuela, and other countries.

Based on the component plots of Figs. 18 and 19, we can verify two main clusters of neurons with reservoirs associated with ISC projects: (i) with sandstone reservoirs and (ii) related to carbonate reservoirs. The group of neurons (i) presents porosity values close to 25, permeabilities

of the order of 1000 mD, depth of about 3500 ft, °API close to 16, and viscosity of 100 cP; however, although group (i) consists mostly of applications in sandstone, when compared with steam injection processes, there is little occurrence of the use of ISC in unconsolidated sandstone. Group (ii) is a particular case, being associated mainly with carbonate reservoirs with lighter °API, higher depths, and low permeabilities, characteristic of geological formations such as Red River A, B, and C (Kumar and Fassihi, 1995; MILLER, 1995). In addition, group (ii) presents a lower density of producing wells and injectors compared to group (i). When compared to other thermal methods, especially steam

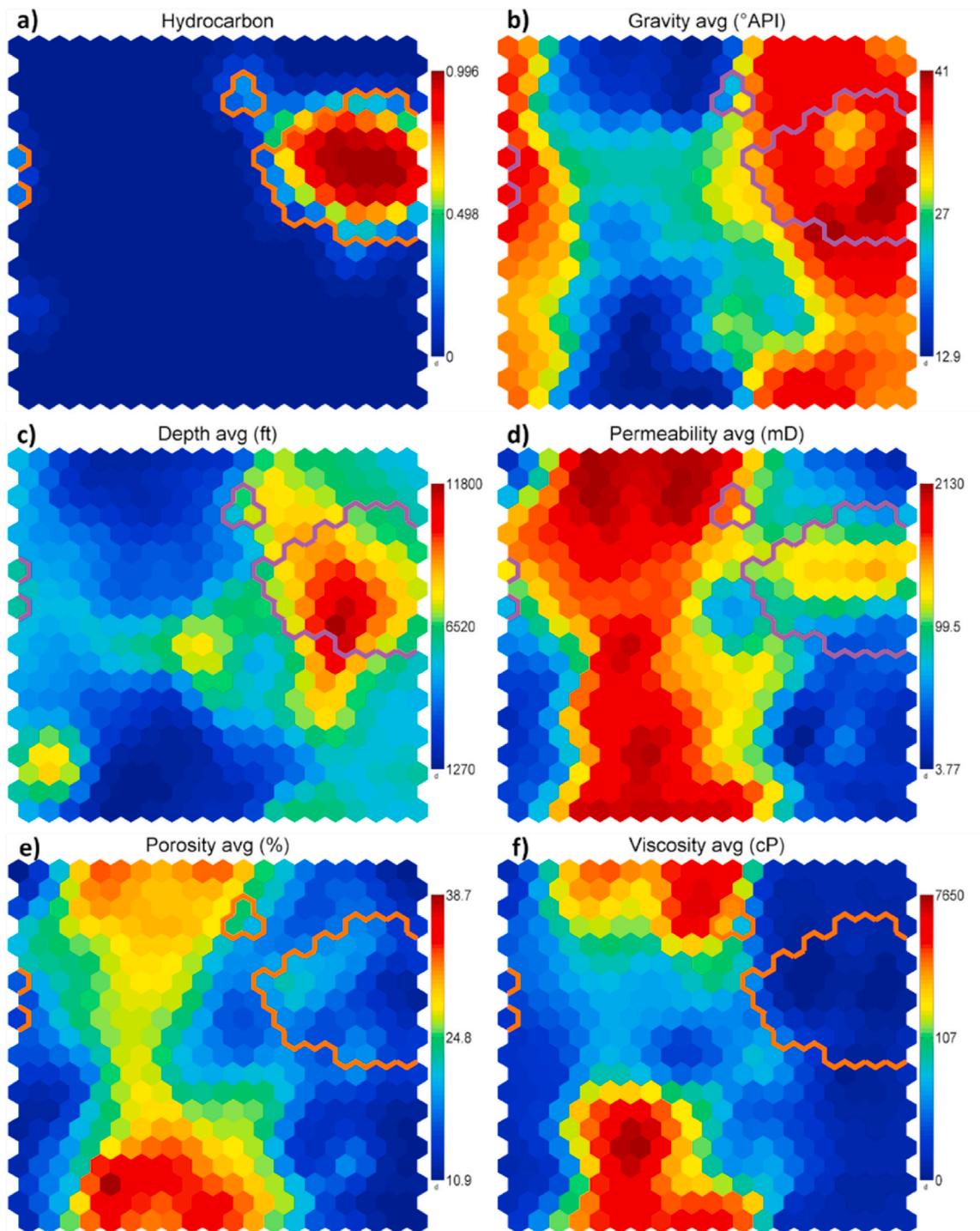


Fig. 20. a) Component plot referring to the method of injection of hydrocarbon gas; b) to f) represent component plots of the other basic variables of the reservoirs. The highlighted line delimits the neurons activated by the hydrocarbon gas injection method.

injection, it is noted that in general, the application of ISC occurs in geological formations with greater depth intervals, satisfying minimum conditions for air injection, in addition to lower permeabilities and porosities, estimated pore throats, and reservoir thicknesses (Kumar and Fassihi, 1995; Miller, 1995).

4.1.4. Gas injection

4.1.4.1. Hydrocarbon. Hydrocarbon gas injection (HC) projects account for 9% of the database, with a total of 142 projects cataloged in 16

countries. Most projects are mainly in the US, Canada, the North Sea region, and China. The first cataloged projects began in the late 1960s in the U.S. and Canada, with wide, field-scale applications slated for the following decades also in North Sea countries.

Based on the component plots of Fig. 20, it can be verified that the mean values for the variables of the cluster sets of higher representativeness of the CO₂ injection projects have mean porosity values of 14%, permeabilities of 90 mD, depth of 9000 ft, temperature of 200 °F, 36 °API, viscosity of 0.8 cP, average estimated pore throat values of 3.4, oil mobility in the order of 200, and adjusted barrel price at the beginning of the project of 45 USD.

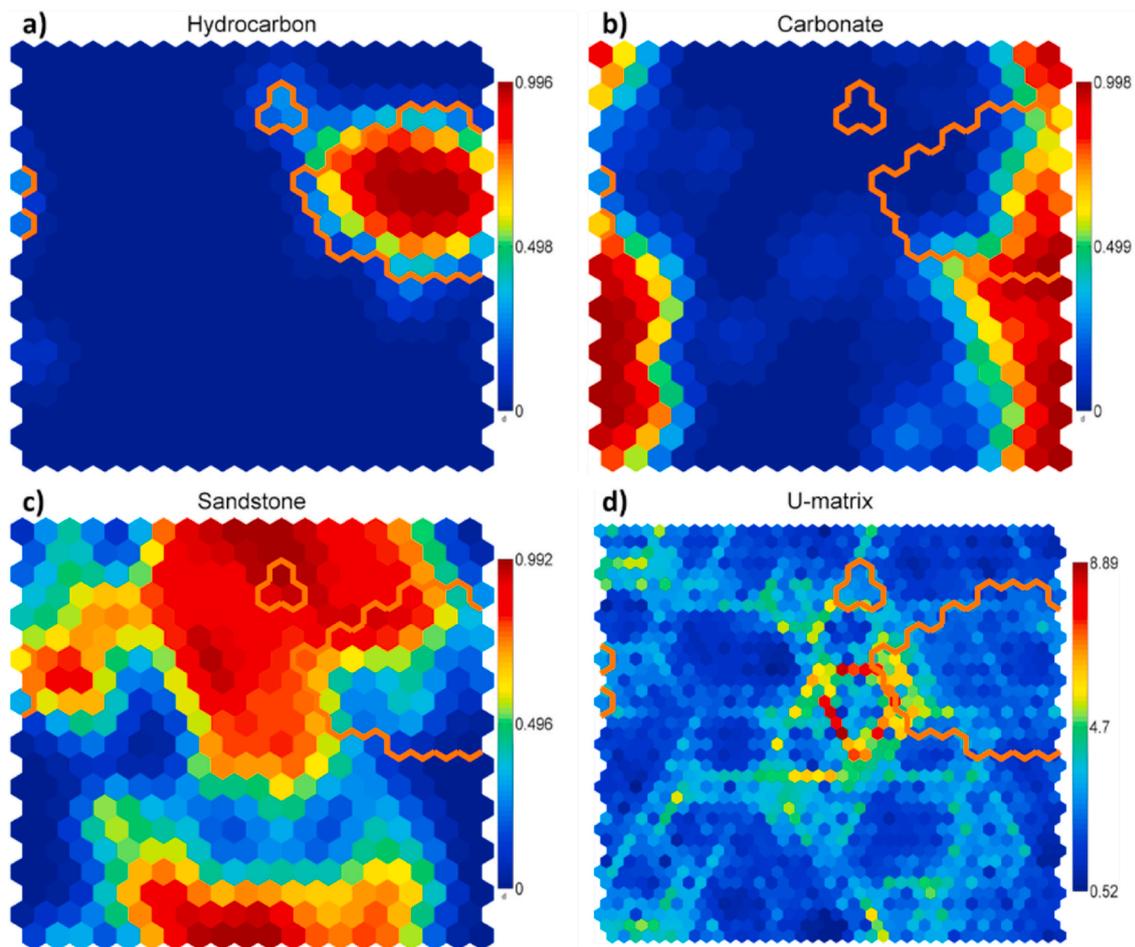


Fig. 21. a) Component plot referring to the injection method of hydrocarbon gas; b) and c) represent component plots of variables related to the lithologies of the reservoirs; d) U-Matrix. The highlighted line delimits the neurons activated by the hydrocarbon gas injection method.

HC injection projects were applied both in sandstone reservoirs and in dolomites and limestone, according to the component plots in Fig. 21. For HC injection projects, when comparing sandstone reservoirs to carbonate reservoirs, sandstone reservoirs have a higher average porosity and permeability and a lower \circ API; for 22 projects, the sandstone reservoirs have depths above 10000 ft. The carbonates that had HC injection have depths of up to 10000 ft except for two projects with depths above 15000 ft. HC injection is the EOR method with the largest offshore number of projects, with 52 offshore projects cataloged, of which 15 are field-scale projects located in the North Sea.

4.1.4.2. Carbon dioxide. CO₂ injection projects account for 32% of the database, with a total of 510 projects cataloged in 15 countries. It is currently the EOR method with the largest number of projects. There is a wide variety of CO₂ injection configurations, with variations concerning the miscibility (if the process is miscible or immiscible), continuity (continuous, alternating with water injection or simultaneous), and the type and concentrations of the fluids injected. Most projects are mainly in the USA, Canada, China, and Brazil, with fewer projects in countries such as Trinidad Tobago, Turkey, North Sea, Argentina, Hungary, Italy, Malaysia, and Qatar. In the USA, CO₂ application began in the 1960s, with wide large-scale application in the following decades. In Canada, Trinidad and Tobago, Brazil, and Turkey, although in smaller numbers, their first CO₂ injection projects were in the 1970s and 1980s, totaling 36 projects in this period for these four countries. In China, a significant number of CO₂ injection projects began to emerge only in 2005. However, by 2010, almost 70% of CO₂ injection projects were implemented in the US; this number has decreased to 40% so far, with the growth of its

application in other countries.

Based on the component plots in Fig. 22, projects with miscible CO₂ processes have a lower depth limit compared with other EOR methods and slightly higher than those referring to immiscible processes. Such minimum depth is necessary to promote miscibility and to remain below the fracture pressure. Miscible projects are also applied at higher depths but with an injection pressure that is not enough to promote miscibility. However, the difference between the miscible and immiscible CO₂ processes stands out with the average \circ API in which they are applied. Projects with immiscible processes feature oil with a wide viscosity range and \circ API ranging from 11 to 40, while projects with miscible processes are concentrated from 27 to 45 \circ API.

According to Fig. 23, the sets of neurons with the highest amount of CO₂ injection projects have mean porosity values of 16%, permeabilities of 30 mD, depth of 5600 ft, temperature of 145 °F, API of 33°, viscosity of 14 cP, estimated high values of pore throat size (between 3 and 5.5) and transmissibility, and an adjusted average price of the barrel at the beginning of the project of 58 USD. CO₂ injection projects were applied both in sandstone and in dolomite and limestone reservoirs, in addition to punctual applications in chalk and chert.

4.1.4.3. Nitrogen. N₂ injection projects represent 2.7% of the database, with a total of 37 projects cataloged in 9 countries. Most projects are mainly in the U.S. and China, with specific projects in Canada, Mexico, Brazil, Trinidad and Tobago, and the Middle East. The first cataloged projects began in the late 1980s in the US. In China, the first reported N₂ injection project was in 2004, with its adoption growing in the country since then.

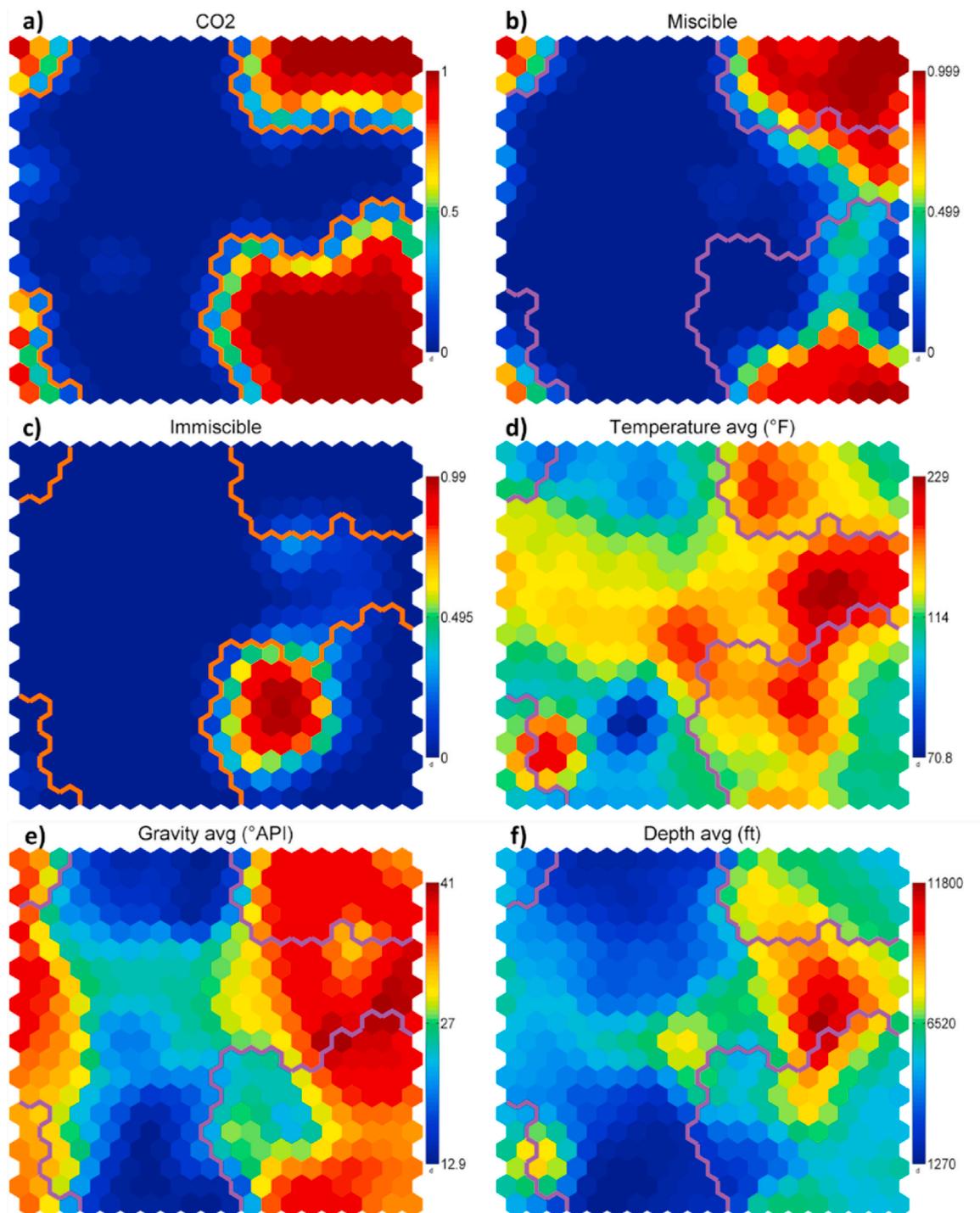


Fig. 22. a) Component plot referring to the CO₂ gas injection method; b) and c) represent component plots related to whether the injection process is miscible or immiscible; d) to f) component plots of basic variables of the reservoirs. The highlighted line delimits neurons activated by the CO₂ gas injection method.

Although the number of projects obtained from N₂ injection is small, the component plots in Fig. 24 show great variability in the application range. This wide range of variability can be explained, in part, by the fact that 27 of the N₂ injection projects use immiscible processes and there were only 5 miscible injection projects. The application of immiscible processes has a lower degree of requirement concerning compliance with requirements such as °API and minimum depths when compared to miscible processes, which aim to promote the necessary conditions for the miscibility of the oil with the injected gas. In particular, it is noted that while the immiscible N₂ injection designs have application at various depths, °API, and oil viscosities, the miscible

designs have application at higher depths and °API and strictly lower viscosities. N₂ injection projects have higher median depths compared to CO₂ projects. In N₂ projects, the average viscosity of the reservoir oil is the highest among the gas methods, possibly associated with the fact that they are projects with a miscible process, being more focused on increasing the reservoir pressure and mechanical displacement of the oil than on the miscibility effects.

Based on the component plots in Fig. 24, it can be verified that the mean values for the variables of the most representative sets of N₂ injection projects have mean porosity values of the order of 15%, permeabilities of 50 mD, depth of 8000 ft, temperature of 200 °F, API of

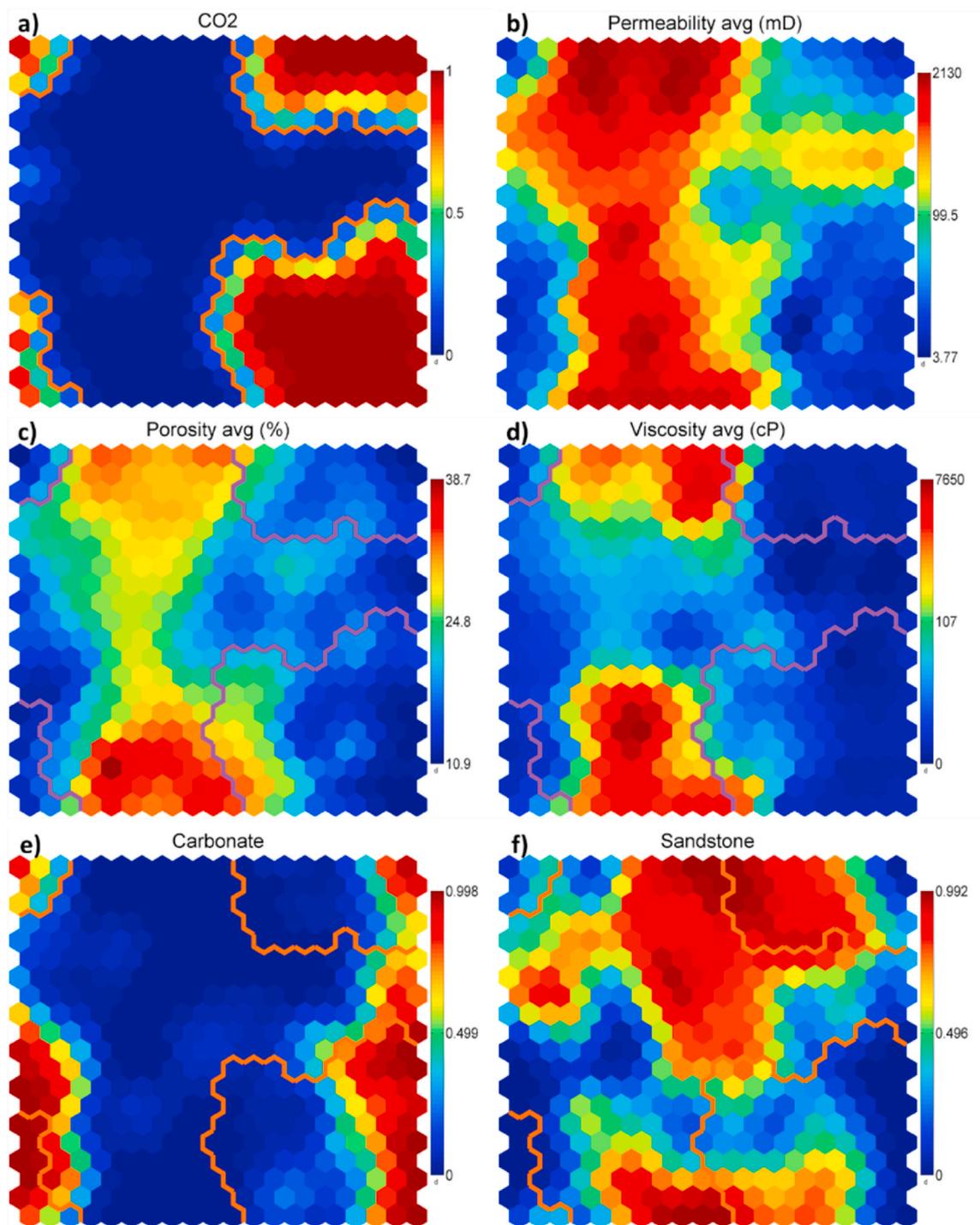


Fig. 23. a) Component plot referring to the CO₂ gas injection method; b) to d) represent component plots of variables related to basic variables of the reservoirs; e) and f) component plots of variables related to the dominant lithologies of the reservoirs. The highlighted line delimits the neurons activated by the CO₂ gas injection method.

38°, viscosity of 10 cP, average estimated pore throat values of 3, oil mobility on the order of 150, and adjusted barrel price at the beginning of the project of 40 USD. According to Fig. 25, N₂ injection projects were applied predominantly in carbonate reservoirs, with only five cases in sandstone reservoirs. N₂ injection projects in carbonates had °API ranging from 30 to 53 and depth ranging from 7500 to 18000 ft. N₂ injection projects in sandstone reservoirs were concentrated in the ranges of 16–25° API and depths of 3000–4600 ft.

According to Fig. 26, among the gas injection methods, N₂ presents a higher average start date for the projects. It also has the lowest average

density of injector wells, with the number of producing wells and low injectors in the range of 4–50 wells, except for two outliers. These two outliers are the onshore fields of Hawkins, Texas, and Elk Hills, California, which have a large number of producing wells and injectors, low depths of around 4600 and 3000 ft, respectively, and which, among other factors, allow N₂ injection patterns with a lower spacing and a lower relative cost.

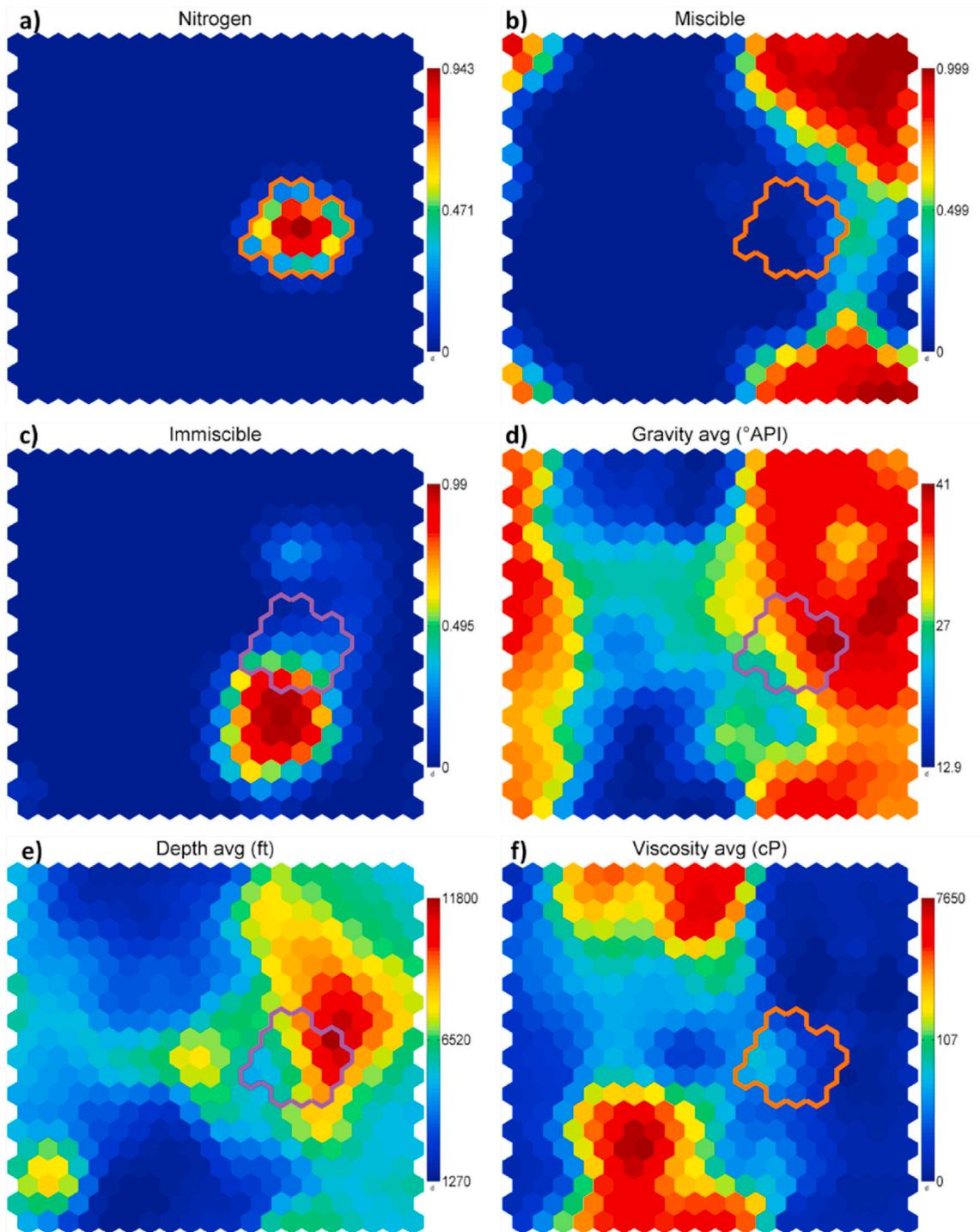


Fig. 24. a) Component plot referring to the nitrogen gas injection method; b) and c) represent whether the gas injection process is miscible or immiscible; d) to f) represent component plots of basic variables of the reservoirs. The highlighted line delimits the neurons activated by the nitrogen gas injection method.

4.1.5. Injection of chemicals

4.1.5.1. Polymer. Polymer injection projects account for 10% of the database, with a total of 168 projects cataloged in 16 countries. Most projects are in China, while other countries such as the USA, Brazil, North Sea countries, Argentina, India, Canada, Venezuela, Colombia, Suriname, and Angola have fewer projects.

Based on the component plots in Fig. 27, it can be verified that the mean values for the variables of cluster sets with the highest number of polymer injection projects have significantly lower variation ranges

than those presented in the boxplots in Fig. 12. More than 90% of BMUs and polymer injection projects have porosity values of 17%–32%, permeabilities from 90 to 1700 mD, depth from 2000 to 7000 ft, temperature from 65 to 175 °F, API from 16 to 33°, and estimated transmissibility above 9000 mD ft cP⁻¹.

As the component plots in Fig. 28, polymer injection is mostly applied to sandstone reservoirs; to a lesser extent, there are also applications in conglomerates. There are few polymer injection projects in carbonates, which are characterized by low matrix permeabilities and the presence of natural fractures (e.g., Leveland fields, Mabee). Polymer

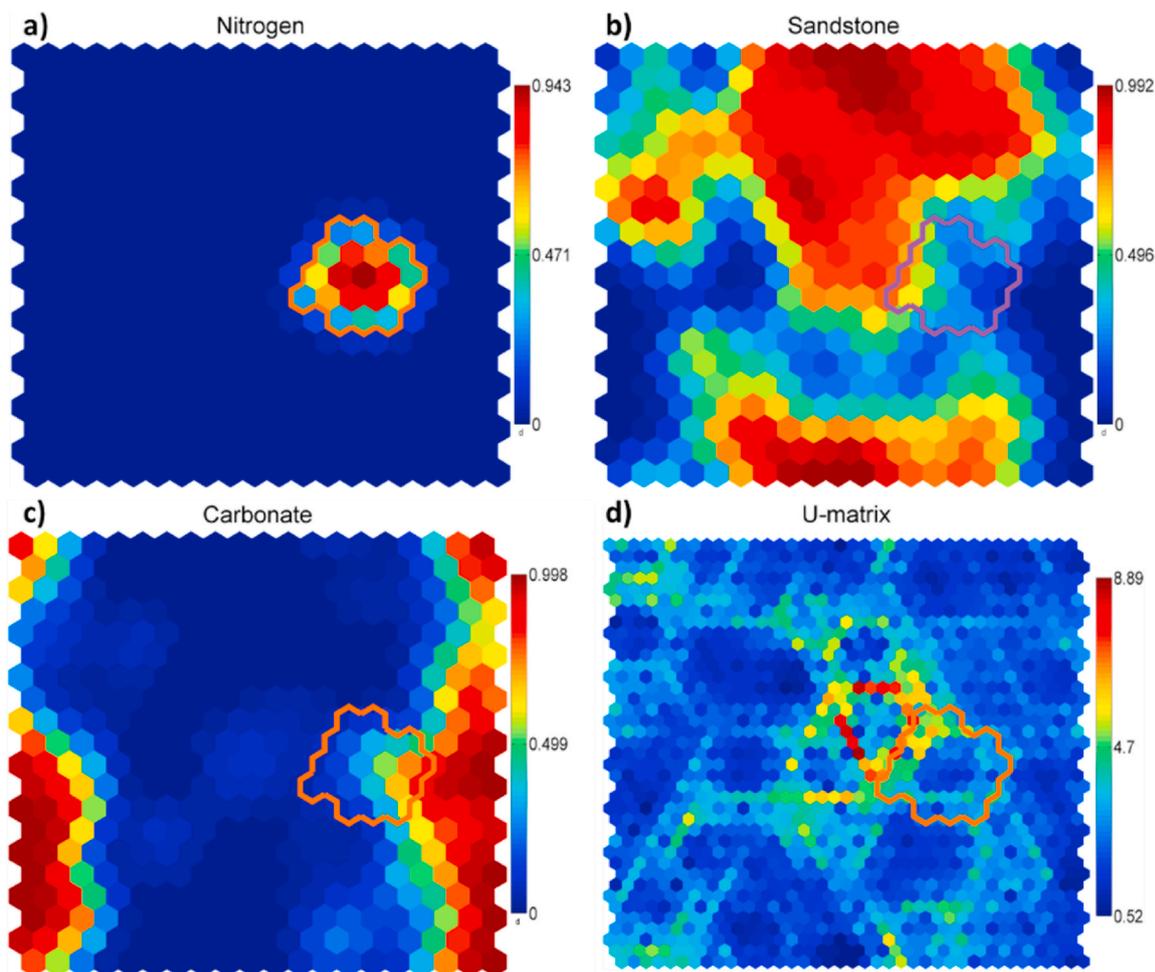


Fig. 25. a) Component plot referring to the projects with nitrogen gas injected; b) and c) represent component plots referring to the dominant lithologies of the reservoirs; d) U-matrix. The highlighted line delimits the neurons activated by the nitrogen gas injection method.

injection usually does not occur at low permeabilities except for carbonate reservoirs: of the 12 projects with permeabilities less than 100 mD, 9 were in carbonates.

Based on the boxplots of Fig. 29, among the chemical EOR methods, polymer injection presents, on average, the highest values of transmissibility and estimated pore throat. Among the chemical methods, polymer injection also has the highest mean values of estimated pore throat, implementation area, and the number of producing wells and injectors and the lowest mean saline concentration (TDS). Concerning TDS, there are outliers with high TDS of the order of 20000, and these recent projects and pilots began their injection after 2016; however, most projects have a TDS of less than 35000 ppm.

Polymer injection is the chemical method with the largest number of offshore projects applied on a field scale, totaling more than 20 fields; looking at the component plots of Fig. 30, in neurons that correspond to the intersection between polymer injection projects and offshore projects, it is noteworthy that this region coincides with an average price of the barrel adjusted for inflation of around 77 USD and that they have average values of porosity and permeability also above the onshore projects. Examples of offshore polymer injection projects include fields located in the Bohai Bay basin in China, Dos Cuadras, The North Sea, and the Campos Basin in Brazil.

The incremental recovery factor associated with polymer injection projects is shown to be in the range of 3%–10%. In most projects, small volumes and concentrations of polymer were used, especially on the pilot scale. In addition, minor changes and improvements are difficult to identify and assign as a response to polymer injection. However, there

are cases of fields with clear successful responses, such as Marmul, Oerrel, Courtenay, and Daqing. These fields have common characteristics, such as permeability averages of the order of 800 mD, low reservoir temperatures of up to 140 °F, high oil saturations at the beginning of the project, and low salinity (TDS), obtaining incremental recovery factors from 11% to 30%.

4.1.5.2. Surfactant and polymer (SP). Surfactant and polymer (SP) injection projects together represent 8% of the database, with a total of 124 projects cataloged in 9 countries. Most projects are mainly in China and the USA, followed by Indonesia, UAE, UK, Hungary, and Malaysia.

Based on the component plots in Fig. 31, it can be verified that the mean values for the variables of cluster sets with the highest number of polymer injection projects have a small range compared to the other EOR methods. More than 90% of clusters and SP injection projects have porosity values of 11%–23%, permeabilities from 10 to 110 mD, depth from 3000 to 6000 ft, temperature from 90 to 140 °F, °API from 28 to 38, viscosity from 1 to 20 cP, and estimated high pore throat values between 2 and 3.5.

SP injection has applications in both dolomite and limestone carbonate reservoirs as well as sandstone. Among the chemical methods, SP is the technology with the highest application in carbonate reservoirs, both in percentage and in absolute number. However, among the chemical methods, SP is the one with the highest percentage of pilot applications, to the detriment of the application on a reservoir scale or commercial scale. Most SP projects have been in place since 2008, with adjusted oil barrel prices above \$70, as shown by the component plots in

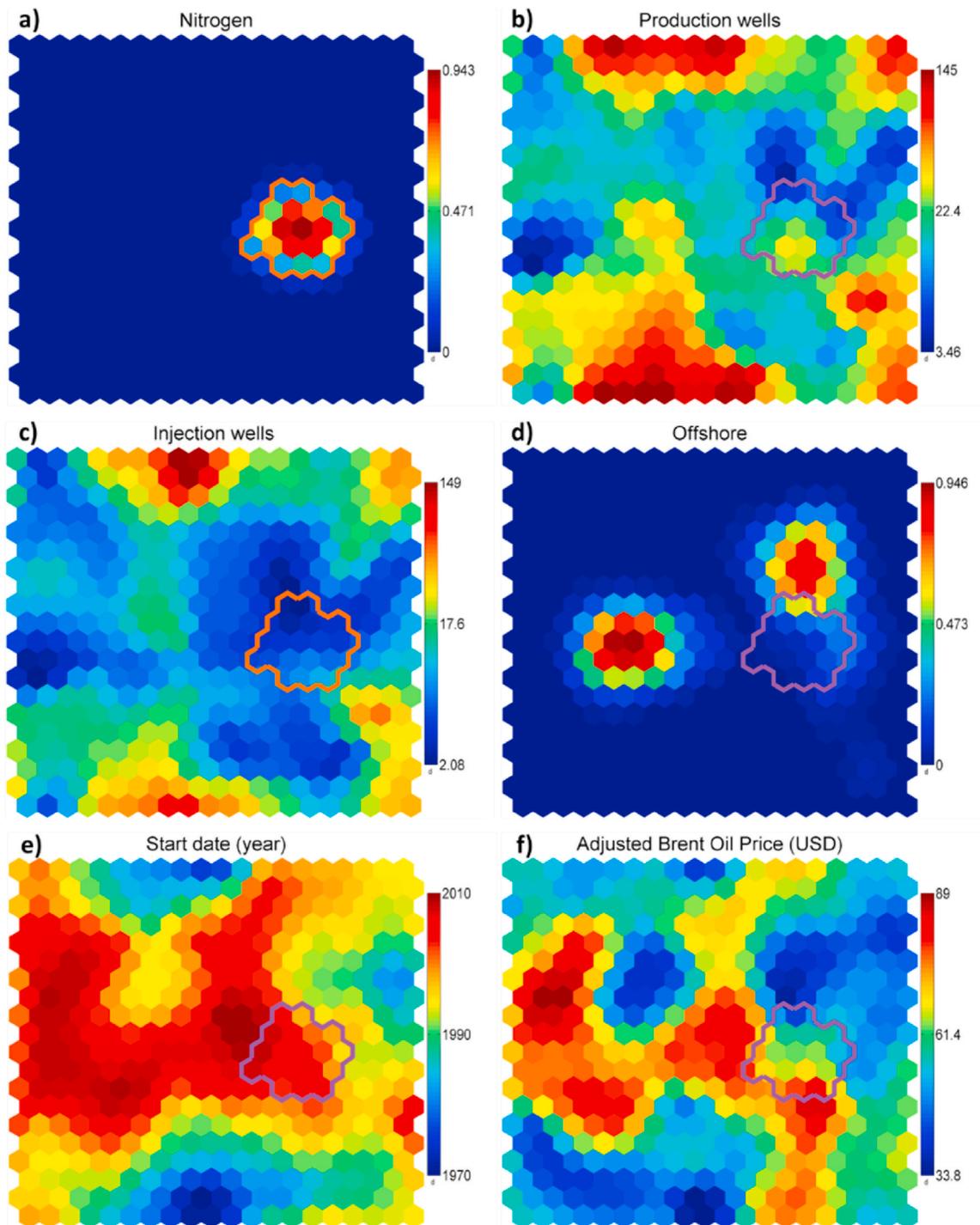


Fig. 26. a) Component plot referring to the nitrogen gas method; b) and c) represent the number of producing wells and injectors; d) whether the project is offshore; e) start year of the EOR projects; f) the price of the barrel adjusted for inflation at the start time of the projects. The highlighted line delimits the neurons activated by the nitrogen gas injection method.

Fig. 32. SP injection has only 5 projects cataloged of offshore applications that have field and commercial scale, among which two are located in Bohai Bay, China, and in fields with an average $^{\circ}$ API equal to 20. Most of the SP injection projects had previously used water injection, with few cases in which it started directly from the primary production. Most SP projects were conducted by NOCs, followed by IOCs.

4.1.6. Other methods

4.1.6.1. Microbial. Microbial injection projects represent 2% of the

database, with a total of 30 projects cataloged in 4 countries. Most projects are mainly in China and the USA, with smaller applications in Russia and Brazil.

Based on the component plots in Fig. 33, it can be verified that the mean values for the variables of the sets of neurons with a greater amount of microbial injection projects have a lower range than the same variables in the boxplots in Fig. 12. More than 90% of neurons and polymer injection projects have porosity values of 14–21, permeabilities from 90 to 700 mD, depth from 2000 to 6000 ft, temperature from 85 to 180 $^{\circ}$ F, API from 25 to 31 $^{\circ}$, viscosity from 5 to 200 cP, estimated pore

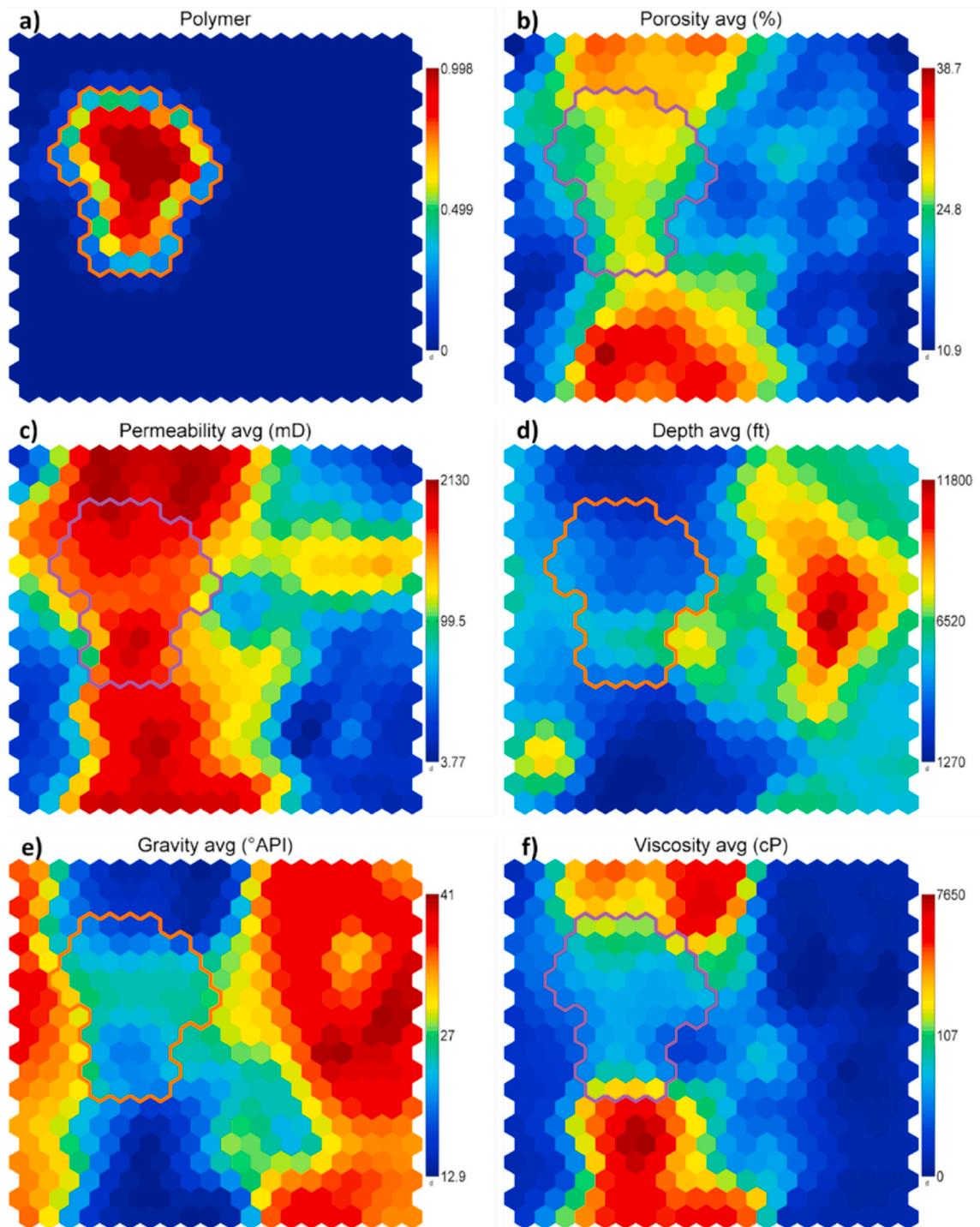


Fig. 27. a) Component plot referring to the polymer flooding method; b) to f) represent component plots of basic reservoir variables. The highlighted line delimits neurons activated by the polymer injection method.

throat values between 2.8 and 4, and oil mobility from 1 to 20.

As the component plots in Fig. 34, the microbial method is mostly applied to sandstone reservoirs; to a lesser extent, there are also applications in conglomerate sandstone. According to Fig. 33, it is possible to observe that microbial injection, when compared to other EOR methods classified as "Other", have lower values of depths and temperatures of the reservoirs in which they are applied. Neurons with the greatest influence on microbial injection projects have low values for the offshore variable due to most microbial projects being in an onshore environment. Among the few offshore projects, we highlight the applications in the Norne fields, in the North Sea with nitrate-reducing bacteria, and

Bokor in Malaysia with the use of cyclic recovery of injected bacteria (AWAN, 2008; KOWALEWSKI et al., 2006; PATEL et al., 2015; RASSENFOSS, 2011; Awan, 2008; Kowalewski et al., 2006; Patel et al., 2015; Rassenfoss, 2011).

In the component plots of Fig. 35, it is also possible to visualize that neurons with a greater influence of microbial projects are related to microbial injection projects with start dates ranging from the 1990s to the present day. Such neurons also show the association of microbial methods mainly at the scale of pilot projects and values or nulls for the field scale variable. Among the few microbial injection projects implemented on a field scale it is Baolige in China (KE et al., 2018), a heavy oil

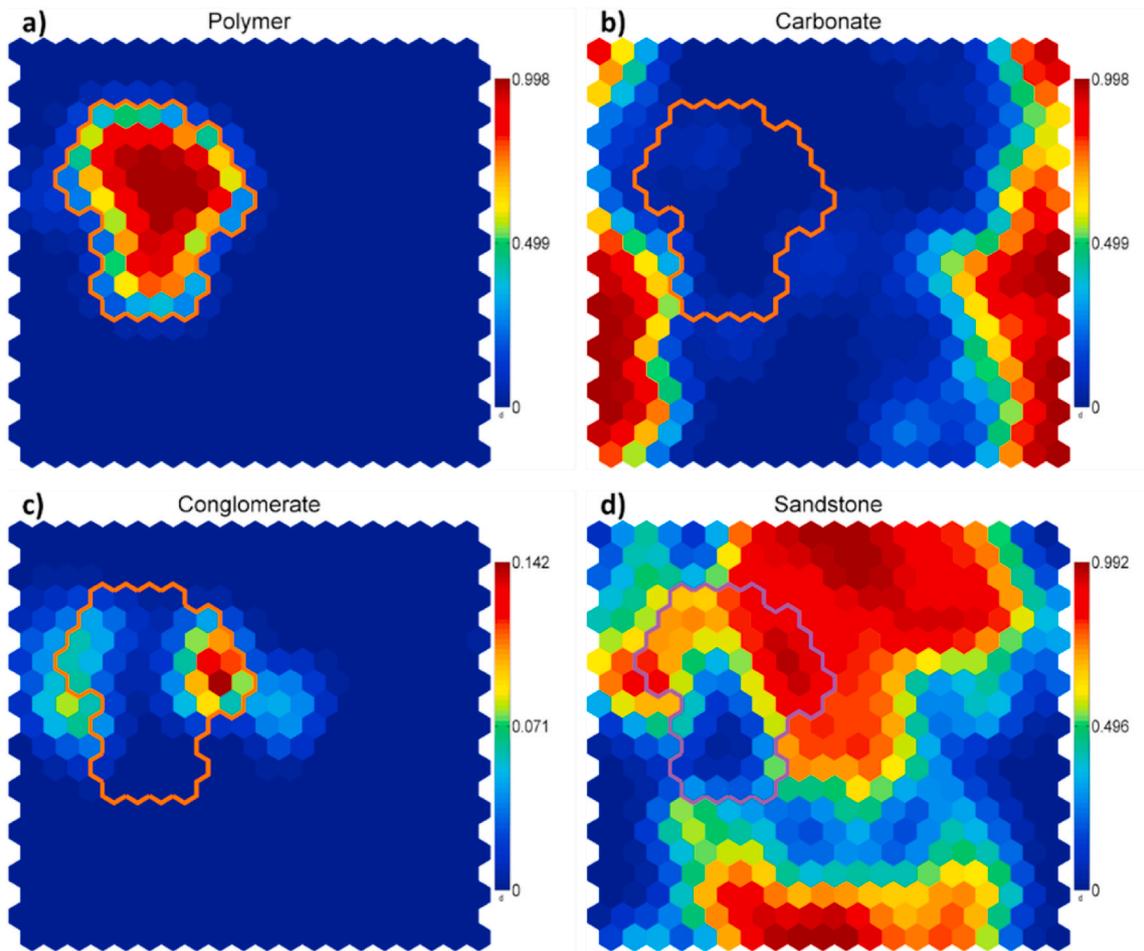


Fig. 28. a) Component plot referring to polymer flooding method; b) to d) represent component plots referring to the dominant lithologies of the reservoirs. The highlighted line delimits neurons activated by the polymer injection method.

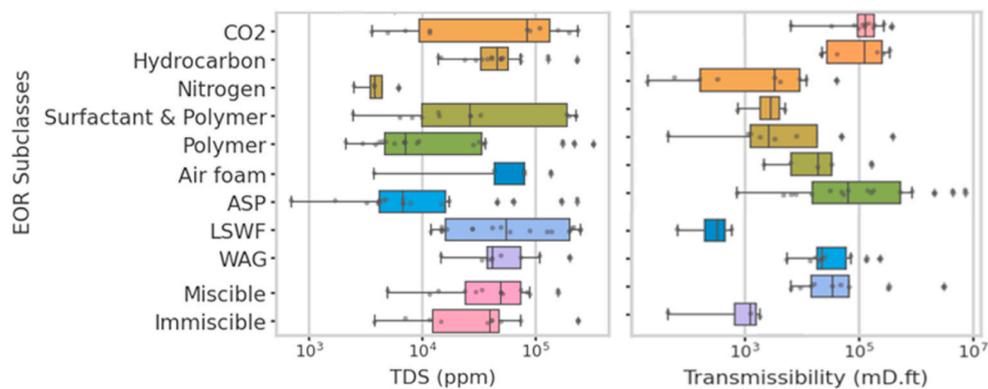


Fig. 29. Boxplots for variables a) TDS and b) transmissibility according to categorized EOR methods.

field with an average viscosity of 157 cP, porosity of 18%, and formation temperature of 122 °F, in which exogenous bacteria and nutrients were injected through 78 injection wells over 40–60 days, followed by water injection; these conditions were favorable for the microbial application, and the project obtained an extra accumulated oil production of around 210000 tons in 43 months. As can be seen in the maps in Fig. 35, microbial injection projects were carried out mainly by large companies, such as NOCs, mainly from the 1990s in China (Ke et al., 2018).

4.1.6.2. Low salinity water (LSWF). LSWF projects represent 2% of the

database, with a total of 33 projects cataloged in 9 countries. Most projects are in the U.S. and the North Sea and Middle East countries.

Based on the component plots in Fig. 36, it can be verified that neurons representing polymer injection projects have porosity values of 14%–25%, permeabilities from 50 to 1000 mD, depth from 4000 to 8000 ft, temperature from 90 to 220 °F, °API from 25 to 32, and viscosity from 0.5 to 40 cP.

According to the component plots of Fig. 37, the neurons that represent the application of LSWF projects are associated with the application predominantly in sandstone reservoirs, with a small

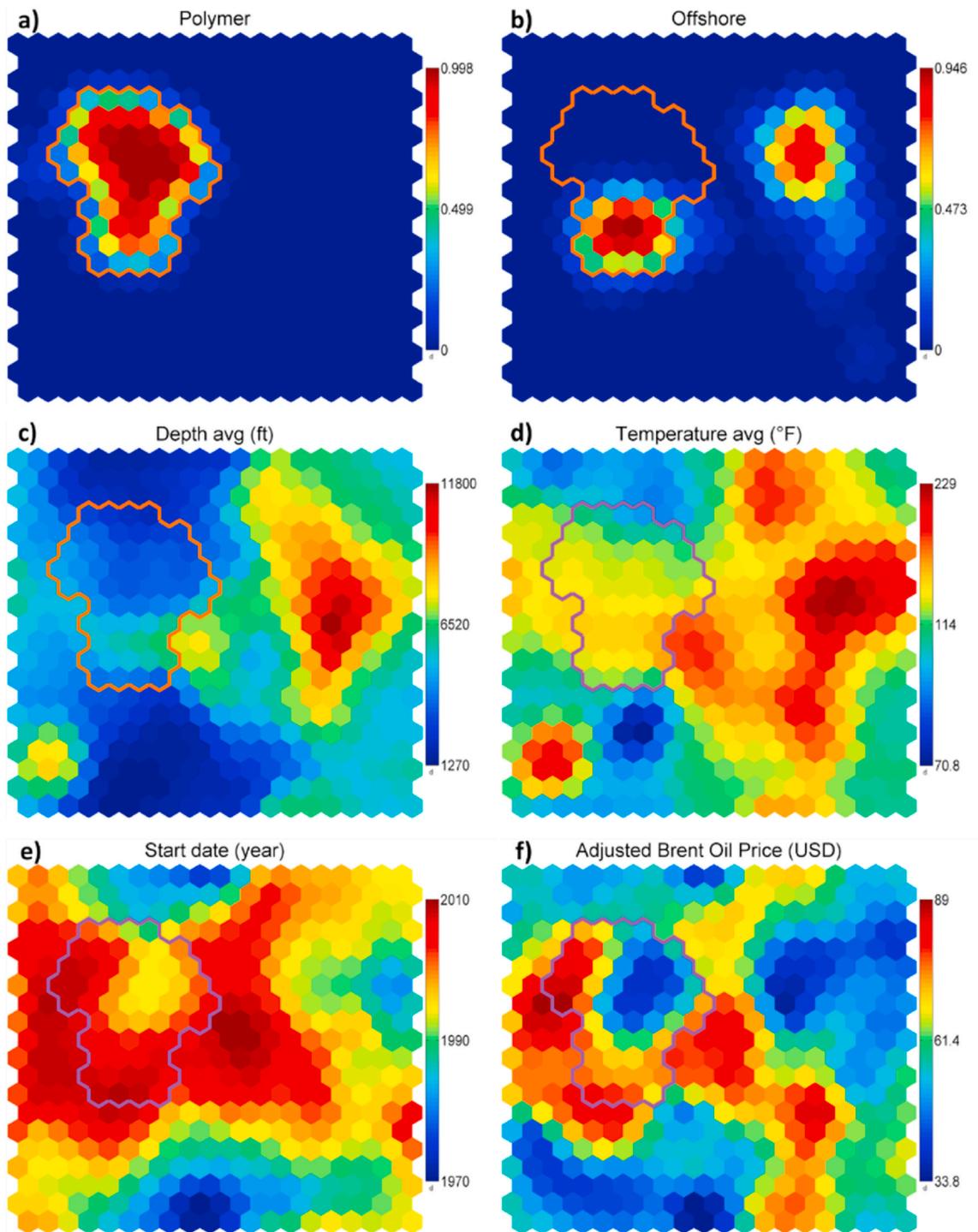


Fig. 30. a) Component plot referring to the polymer flooding method; b) whether the project is offshore; c) to e) the depth and temperature of the respective reservoirs and their start dates; f) inflation-adjusted barrel prices at the beginning of the project. The highlighted line delimits neurons activated by the polymer injection method.

presence of applications in carbonates. According to Fig. 38, the use of LSWF on a field scale began in the 2000s, with growth in its use from 2010, a period with the high price of the barrel. Despite its recent application, LSWF stands out for already presenting field applications in offshore environments, such as the Endicott sandstone fields in Alaska and Snorre, Heidrun, and Clair Ridge in the North Sea (CHEN et al., 2021). The fields with LSWF projects were mostly conducted by IOCs and outnumbered by NOCs, usually with a nearby source of water catchment (such as seawater), and with no cataloged applications

carried out by small and medium-sized independent operators (Chen et al., 2021).

5. Discussions

The EOR methods are strategic tools and have great importance for the full development of oil reservoir production. The results presented visually summarize information already supported by the literature regarding reservoir engineering. In addition, the results made

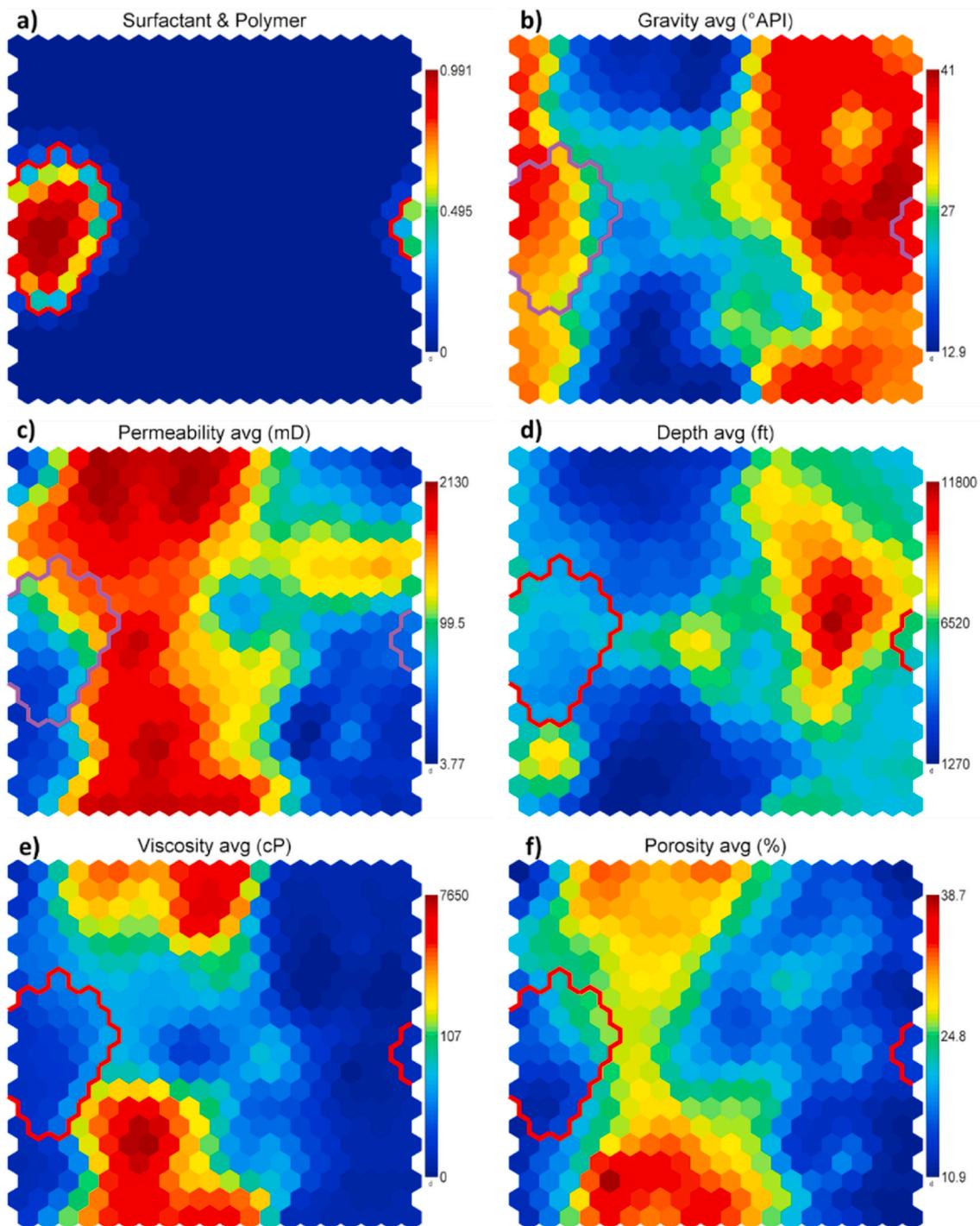


Fig. 31. a) Component plot referring to the surfactant and polymer injection method; b) to f) component plots of the other basic variables of the reservoirs. The highlighted line delimits the neurons activated by the surfactant and polymer injection method.

it possible to identify patterns in other technical, economic, and also strategic aspects of the application of the companies. This approach denotes as relevant the application of data science techniques for the analysis of EOR projects, both for the observation of similar reservoirs and analysis related to the behavior of variables in these reservoirs. Some discussions related to the methods adopted as well as the results obtained will be addressed below.

Proposed analysis methodology

The results presented showed cohesion between patterns observed

from the SOM maps for the real and theoretical models related to EOR. In addition, the information obtained demonstrated that the integration and treatment of the different data sources were performed efficiently, allowing a broad and cohesive dataset about EOR methods used on a global scale to be constituted. The use of synthetic attributes, such as estimated pore throat size and oil mobility, allowed classes of EOR methods to be distinguished and a more in-depth analysis of technical parameters and their implications to be achieved. In addition, with the use of data visualization techniques, it was possible to analyze the dynamics of EOR methods in an integrated way, allowing us to better understand the applications, advantages, and technical limitations of

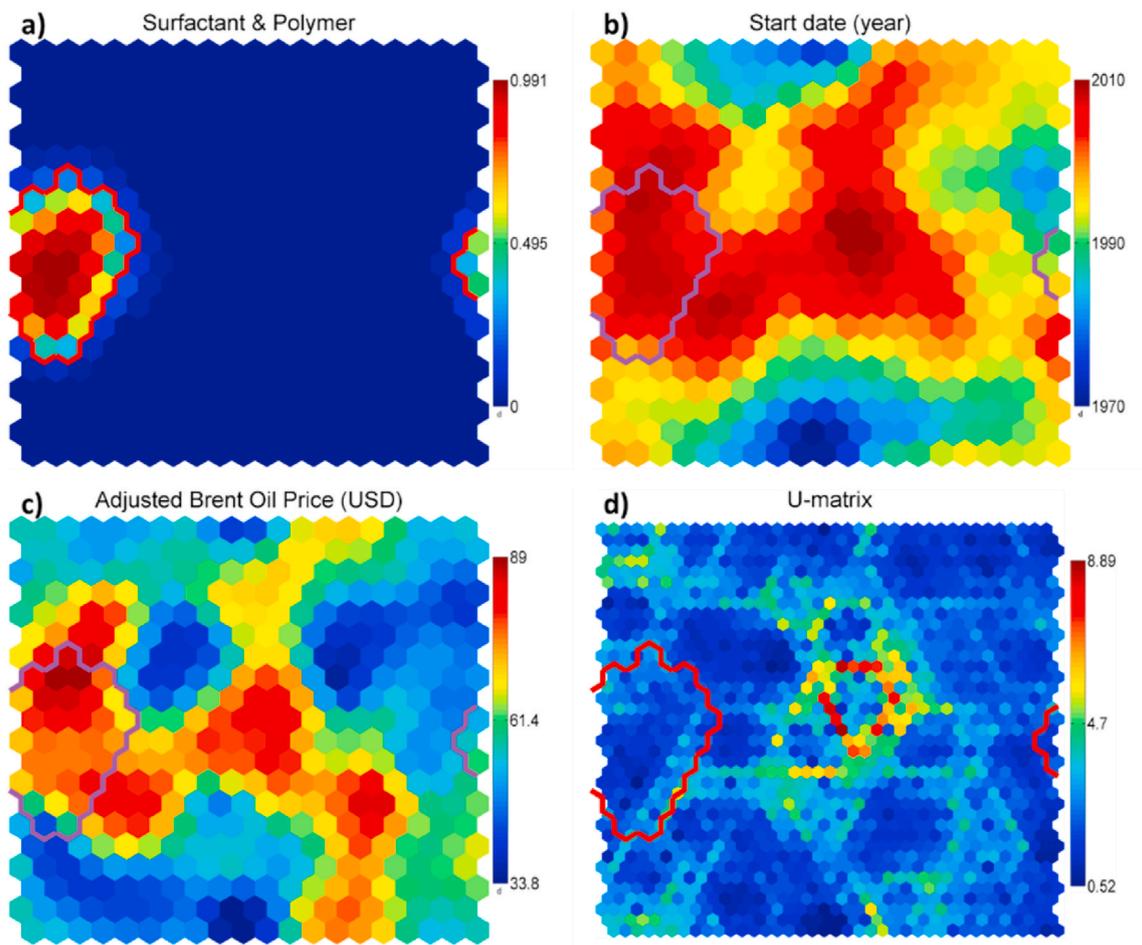


Fig. 32. a) Component plots of the type of injected chemical (surfactant and polymer), b) and c) start date and adjusted price of the barrel in the start year of the projects; d) U-matrix. The highlighted line delimits the neurons activated by the surfactant and polymer injection method.

each EOR method and to compare them with each other.

The limitations of the methodology adopted, however, included aspects such as the difficulty in searching, acquiring, collecting, integrating, cleaning, and checking data besides the occurrence of missing data for several of the variables, which often reduces the sample space of EOR projects within a given analysis. To mitigate such problems, several sources were used that allowed complementarity with the variables and data present and automated pre-processing steps of the data reduced the time spent in this process. In addition, there are limitations in the use of SOMs both in the graphical representation of component plots for each neuron being assigned a value statistically close to that of the projects represented and by smoothing extreme values. As an example, it was not possible to visualize the reservoir maps with oils of °API below 11 or above 40 and permeabilities below 5 mD or 2000 mD, and generalizations were made based on the history and patterns observed. For this, data analysis was performed in conjunction with other visualizations and inspection of the data and validation in the literature. Furthermore, the scope of this work also does not consider in the analyzes aspects related to the economic viability of EOR projects and in which cases they were able to overcome break-even, or possible root causes of failure of certain projects.

The results and patterns identified

5.1.1. Influence of technical variables on EOR methods

5.1.1.1. Basic properties of reservoirs. The analysis of the basic properties of the reservoirs indicates the influence of these parameters on the

choice and application of EOR methods. The development of synthetic parameters, such as estimated pore throat size and oil mobility index, allowed additional patterns to be revealed, bringing a gain of information and evidencing the impact that it can present in the application of EOR. We identified the results of a broad consensus in the literature (Dong et al., 2019; Liu et al., 2020; Mokheimer et al., 2019), as the application of thermal methods of vapor injection in reservoirs was related to (i) heavy oils, to heat the in-situ formation with fluid expansion and increase oil fluidity, decrease residual saturation, promote water wettability; (ii) sandstone formations and high porosities and permeabilities, to have a minimum oil flow capacity; (iii) low depths, so that there is a lower loss in the transmission of steam; (iv) a small number of offshore projects; (v) high density of producing wells and injectors, also made possible by the low depths and to have greater sweep efficiency, among others. Different patterns were also identified for the vapor injection subtypes, with SAGD being applied in extra-heavy oil reservoirs with less thick or carbonate formations and in fields that have often presented other recovery methods. For the application of ISC, similarities were confirmed concerning steam injection processes, such as application in generally viscous oils (Dong et al., 2019; Hashemi-Kiasari et al., 2014; Liu et al., 2020; Mokheimer et al., 2019; Yuan and Wood, 2018; Yuan and Wang, 2018). Differences were observed in terms of application in conditions where vapor injection processes are not economically viable, such as a greater range of reservoir depth, lower porosities and permeabilities, and lithological variation (Ji-hong et al., 2020; Miller, 1995; Mokheimer et al., 2019; Turta et al., 2007).

Regarding gas injection methods, it was found that they are applied

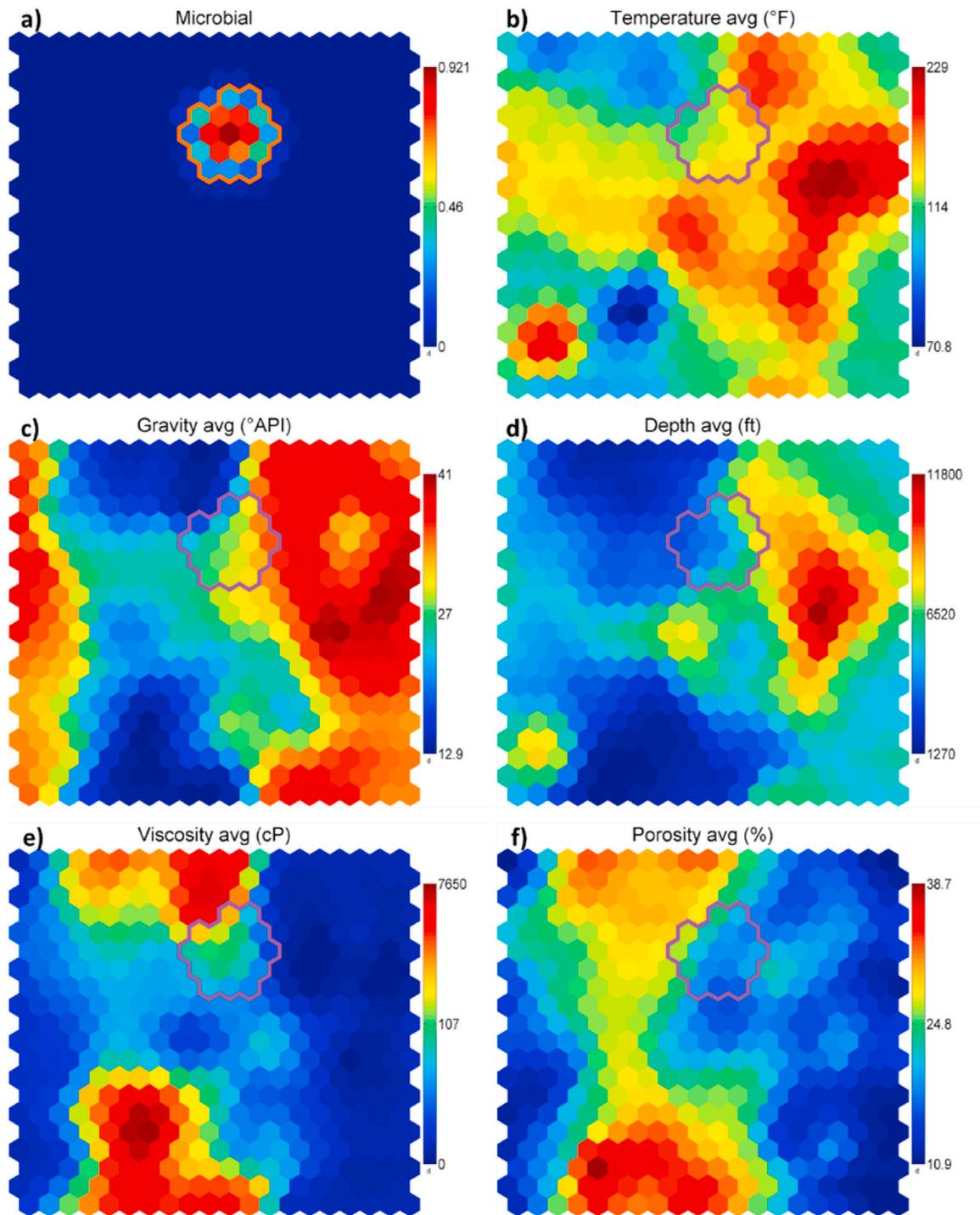


Fig. 33. Component plots referring to a) microbial injection and b) to f) variables of basic properties of the reservoirs. The highlighted line delimits the neurons activated by the microbial injection method.

in reservoirs with general characteristics of (i) low porosity and permeability values, as a way to avoid gravitational segregation of the gas; (ii) low to intermediate viscosities; (iii) average values at high depths and temperatures, to provide minimum conditions for pressure maintenance in the reservoir without fracturing the formation. These patterns are corroborated by literature already established in the area. Several distinctions within this group could be noted both regarding the type of gas injected, as well as concerning the miscibility and continuity or alternation of the injection process, also addressed by work in the area (Alvarado and Manrique, 2010a; Awan, 2008; Belazreg and Mahmood, 2020; J. A. J.A. Brodie et al., 2012; Gbadamosi et al., 2018a; Jin, 2017,

p. 2; Kokal and Al-Kaabi, 2010; Li et al., 2019; Liu et al., 2020; Taber et al., 1997).

For the use of chemical methods, the main patterns of the basic characteristics of the reservoirs were intermediate values of porosities, permeabilities, °API, and viscosities; low values at medium depths and temperatures of reservoirs, and dominant lithology sanders with fewer applications in carbonates. Several particular characteristics of the application of different chemicals were also observed and are validated by the literature (Belhaj et al., 2020, 2020, 2020; Delamaide and Degre, nd.; Gbadamosi et al., 2019a; Sheng, 2010). The EOR methods classified as "Other", which encompass mostly recent techniques, have great

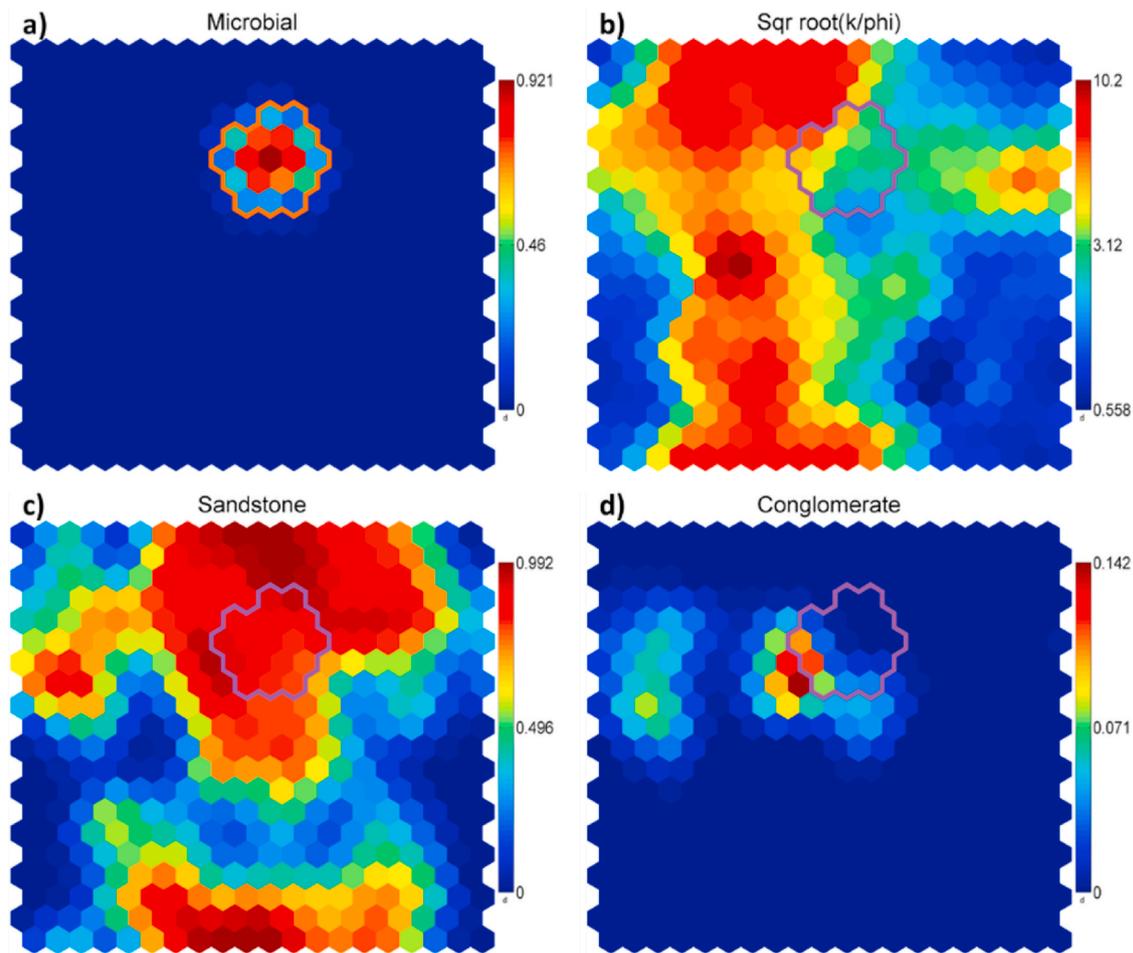


Fig. 34. Component plots representing a) microbial injection and b) to d) variables referring to the dominant lithologies of the reservoirs. The highlighted line delimits the neurons activated by the microbial injection method.

dispersion and low concentration of their data.

5.1.1.2. Offshore environment and field development. The EOR method with the highest number of offshore projects cataloged was the injection of hydrocarbon gas (miscible or immiscible), followed by polymer injection and CO₂ injection. The first offshore EOR field-scale projects took place mainly in the North Sea, associated with neighboring fields with high gas production available for injection. The offshore gas injection has already gained high technological maturity, with several projects around the world.

The application of offshore chemical methods has begun more recently and may be a viable option mainly where there is no availability of gas source for injection. Examples of offshore polymer injection projects include the Dos Cuadras fields in the Gulf of Mexico, Bohai Bay in China, Captain in the North Sea, and Dalia in Angola. Despite successful offshore cases of chemical methods, it is verified that challenges persist related to factors such as (i) high depths, temperatures, and salinities, which can lead to degradation and loss of effectiveness of polymers and surfactants; (ii) large spacing between wells, requiring a higher volume and consequently higher cost of injected chemicals; (iii) logistical difficulties and platform space for the transport and solubilization of large volumes of solutions; (iv) problems in water/oil separation due to low interfacial tension and high viscosity of fluids produced with remaining amounts of injected fluids, resulting in difficulties in the flotation process and limiting coalescence. In this context, polymer injection has a greater advance in offshore applications when compared to ASP/SP because of lower chemical consumption, higher maturity, and lower implementation complexity. On the other hand, there are several

advances in research for the development of new substances and additives that show the possibility of overcoming some of these barriers to the use of chemicals in an offshore (Bondor, 2005; Kang et al., 2016; Raney et al., 2012; Reksidler et al., 2015). Compared to the other classes of EOR methods, steam injection presents a small number of offshore projects due to factors such as heat loss during injection, a lower density of wells to obtain an efficient sweep, challenges related to equipment required for heating, pressurization, and steam injection, and that is difficult to operationalize in an offshore environment, among others.

The thermal methods presented a higher density of wells, and that can be related both to the fact that in reservoirs that have their application, the reservoirs are less deep and the wells present a lower cost, as well as due to the difficulty of viscous oil flow, making it necessary to use infill drilling techniques aimed at avoiding dead oil zones. For gas injection, higher depths are observed, which have a cost impact on good construction and consequently lower well density. For chemical methods, a density of wells higher than that of gas but lower than thermal methods is noted; this can be attributed to reasons such as (i) a very large spacing between wells demands a higher time of the chemical injection in in-situ conditions and its greater susceptibility to degradation; (ii) larger spacings between wells require injection of a larger volume of chemicals; and (iii) steam injection happens to have a higher density of wells due to the lower depth of their reservoirs and at smaller depths, which it is, in general, a more technically and economically viable option in such cases, among other factors.

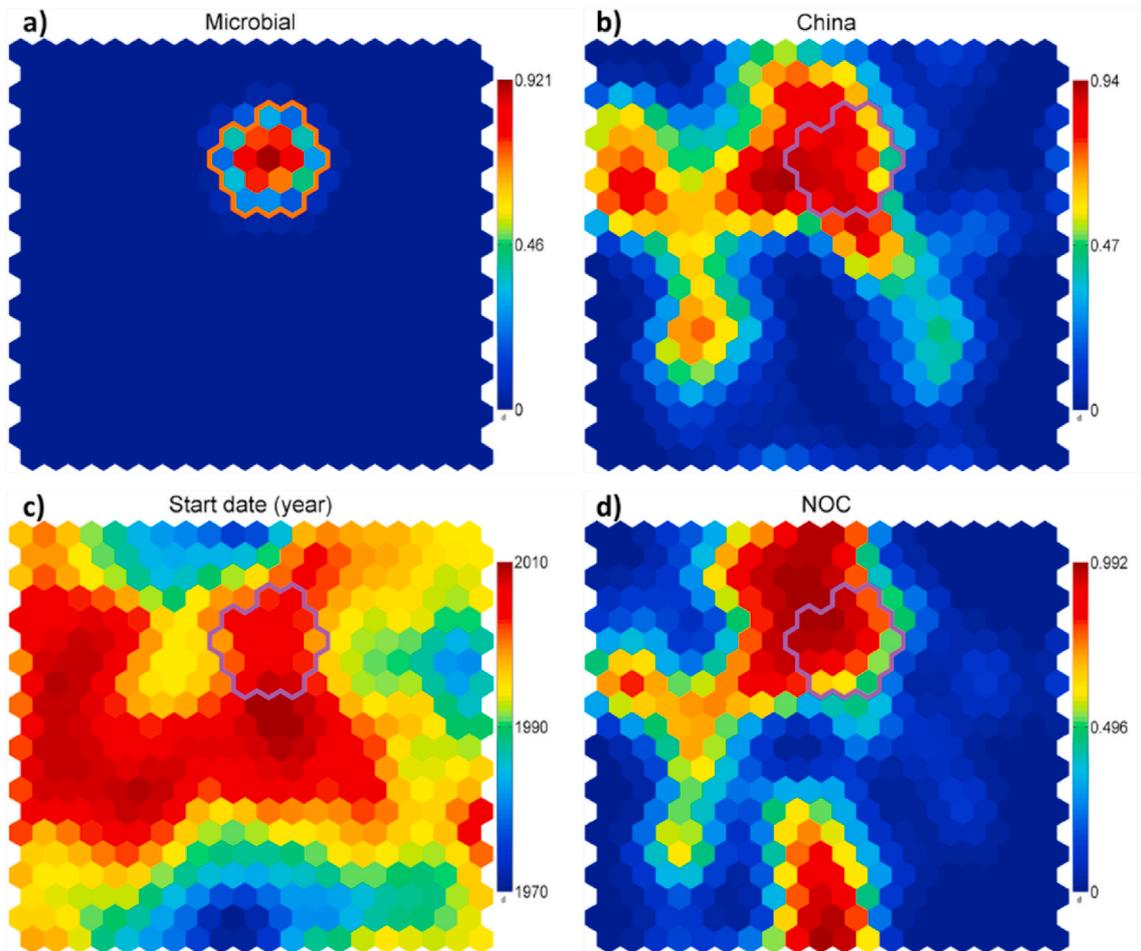


Fig. 35. Component plots refer to) microbial injection; b) implementation of projects in China, c) start year of the projects; d) whether the operating company of each reservoir is classified as the National Oil Company.

5.1.2. Localization influence and non-technical variables in the choice of EOR methods

From the results, it is also possible to observe the influence of aspects related to localization in the choice, application, and economic development of EOR techniques, even when satisfactory technical criteria were presented for the application of EOR. Although the adoption of advanced recovery strategies has spread worldwide, the concentration of the application of certain EOR methods in a small number of localities is noted. These concentrations were shown to be associated with factors such as availability and costs of supplying and transporting injection fluid, the production history of the fields, expertise in the development of technologies by companies, the degree of development of industry and infrastructure, regulatory aspects, and policies to encourage the production of remaining natural resources of each country.

The injection of hydrocarbon gas is one of the EOR methods that has applications in the largest number of countries and regions of the world. On the one hand, for certain cases, studies show the injection of hydrocarbons presents generally lower recovery efficiency when compared to other methods of gas injection, such as CO₂ injection (Gbadamosi et al., 2018a). However, on the other hand, it is noted that many companies decide to apply hydrocarbon injection, to the detriment of other EOR methods that may be more efficient. Among the reasons for this, some factors not exclusively related to the properties of the reservoirs stand out, such as the general availability of gas for injection obtained from neighboring fields, greater knowledge of the behavior and dynamics of the gas in the reservoir, and lower costs associated with corrosion, among others.

Methods such as CO₂ injection and the use of chemicals (polymer and

surfactant) have their application concentrated in a smaller number of countries, such as the USA and China, being associated with factors such as a network of CO₂ transport pipelines and a chemical industry developed with the supply of fine chemicals at low cost, respectively. In addition, policies to encourage the capture and storage of CO₂ have proven to be a great driver for the application of this technology. Thermal methods, in turn, are generally associated with localities that have, among other factors, reservoirs with specific characteristics, such as heavy oil and high pore flow cover. In addition, regulatory policies and tax incentives are important for the continuity of the development of mature or marginal fields and for the application of EOR to give survival and faster production of such assets.

The results also suggest that other variables, not necessarily of a technical nature, can also influence decision-making regarding the implementation of EOR projects and the development of research and recovery technologies. Among these variables, se observa destaca (i) price of a barrel of oil and world economic scenario, (ii) size, segment, and positioning of operating companies, (iii) environmental restrictions, (iv) existence of depleted reserves and mature fields, (v) incentive policies, and (vi) regulatory and political aspects of the countries in which oil reserves are located. Such dimensions may be the object of more detailed study of other work in this area.

6. Conclusions

The methodological routine developed and applied here integrates several modalities of information, processing, and visualization of data, allowing us to explore a wide set of experiences of previous EOR projects

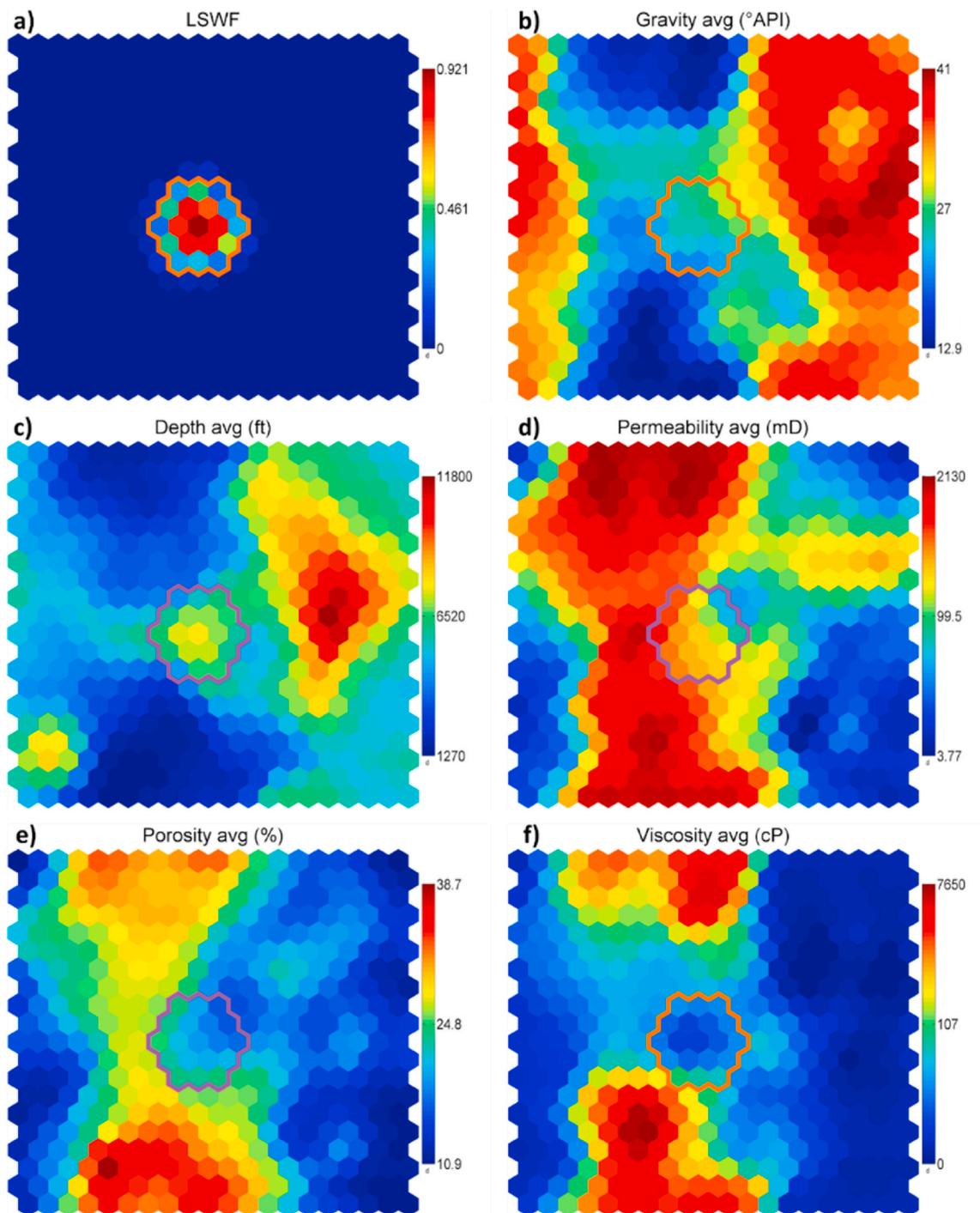


Fig. 36. Component plots related to a) the low salinity water injection method (LSWF); (b) to (f) represent basic properties of the reservoirs.

on a global scale. This approach allowed obtaining a high-level representation of self-contained knowledge of the EOR projects. A comprehensive view of the application of EOR in the world was obtained and it was possible to understand implicit interrelationships between several variables with the different EOR methods, allowing the identification of trends, new directions of research and better informed decisions. In comparison with similar previous works on global or regional surveys of EOR projects, it is noted that it was possible to carry out a detailed analysis, with a representative quantity of the current status of implementation of EOR projects in the world, in addition to addressing petrophysics, location environment, operational and economic parameters.

Based on the results obtained, it was possible to notice that variables

of different natures demonstrate an impact on the choice, application, and technical, economic, and technological success of EOR methods. These variables indicate having impacts that vary according not only to the technical properties of the reservoirs and the EOR method in question but also to local, commercial, and time aspects. This analysis allows us to clarify part of the complex and intricate network of factors that influence the application of EOR methods on a global scale. Its main contribution is also the graphical analysis of variables of different natures that demonstrates an impact on the choice, application, and development of EOR methods.

Moreover, despite the diffusion of secondary and tertiary recovery technologies, there is still a concentration of different EOR methods in a

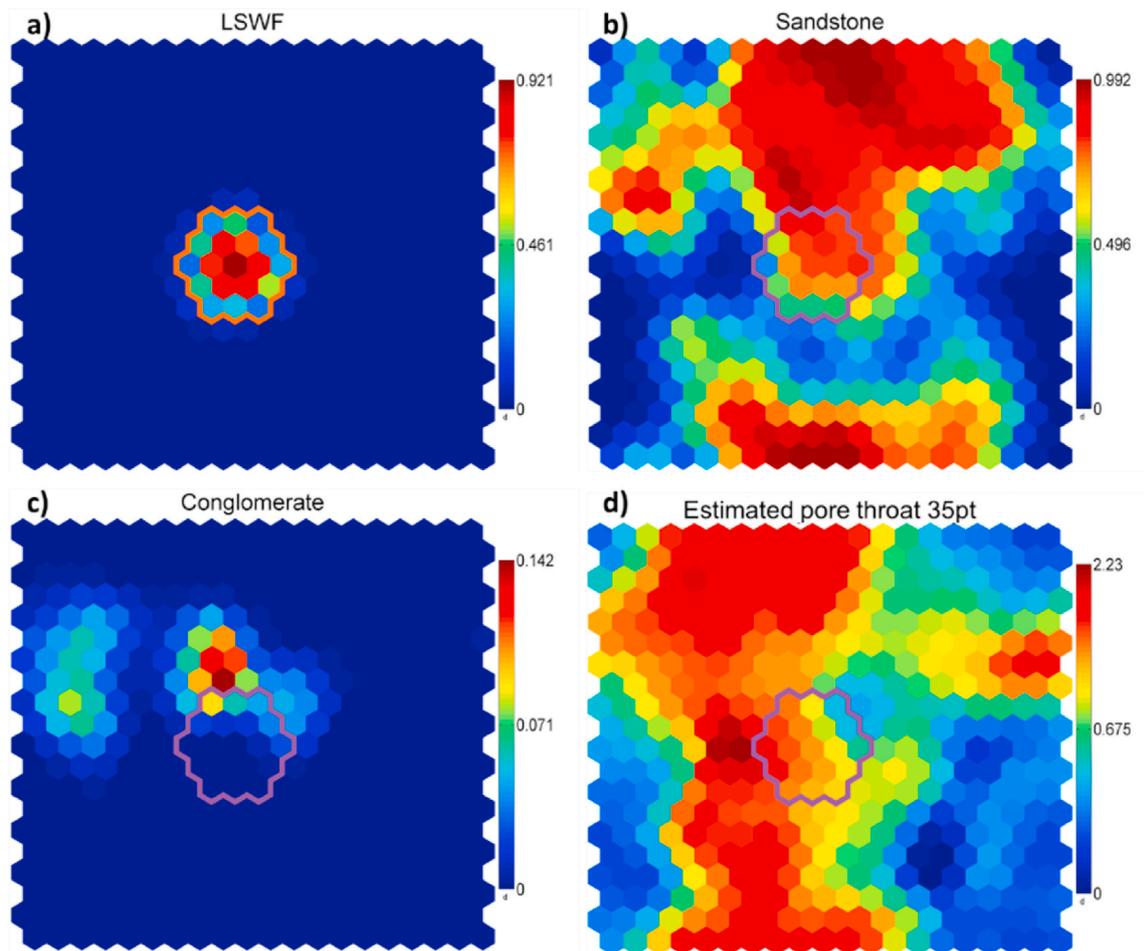


Fig. 37. Component plots related to a) the low salinity water injection method (LSWF); (b) to (d) represent dominant lithologies of the reservoirs.

few regions of the world. Factors were identified for the implementation, maintenance, and expansion of EOR projects ranging from technical reasons, the existence of deselected reserves, available infrastructure and supply and availability of supplies, to aspects related to incentive policies and regulatory aspects of each country, structure, and strategic positioning of the business segments and the world's economic scenario.

It was observed that technical factors related to lithological and flow properties of reservoirs, such as porosity, permeability, depth, temperature, viscosity, ϕ API, salinity of formation, and heterogeneities, among others have different impacts on the choice of different EOR methods. The creation of synthetic variables was also a new approach in this type of work and which allowed for the identification of previously implicit correlations and patterns and deepened the analysis of technical variables and their impacts on EOR methods.

The influence of non-petrophysical variables on the application of EOR methods was also observed. Aspects related to the availability and costs of purchasing and transporting injection fluids have been shown to have a great influence on decision-making processes regarding the choice of the EOR method to be applied. As an example, we can highlight the injection of gas obtained from neighboring fields, the injection of CO₂ with supply from the capture network of specific industries and pipelines, and the injection of chemicals in markets with developed and low-cost industries.

There was the influence of the type of environment in the field in question, whether offshore or onshore, in which technical and economic viability is still challenging. In addition, EOR projects still represent a

small portion of world production but have great growth potential. There is also a diversity of strategic positions concerning the preference of EOR techniques according to the profile of an operational company. The impact and influence of non-technical variables on the adoption of EOR and furthermore, the use of advanced text processing techniques using Natural Language Processing and Large Language Models may be the object of further future studies.

CRediT authorship contribution statement

Luiz Felipe Niedermaier Custodio: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Cleyton de Carvalho Carneiro:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

Luiz Felipe Niedermaier Custodio reports financial support was provided by Petrobras. Luiz Felipe Niedermaier Custodio reports financial support was provided by National Agency of Oil Natural Gas and Biofuels (ANP).

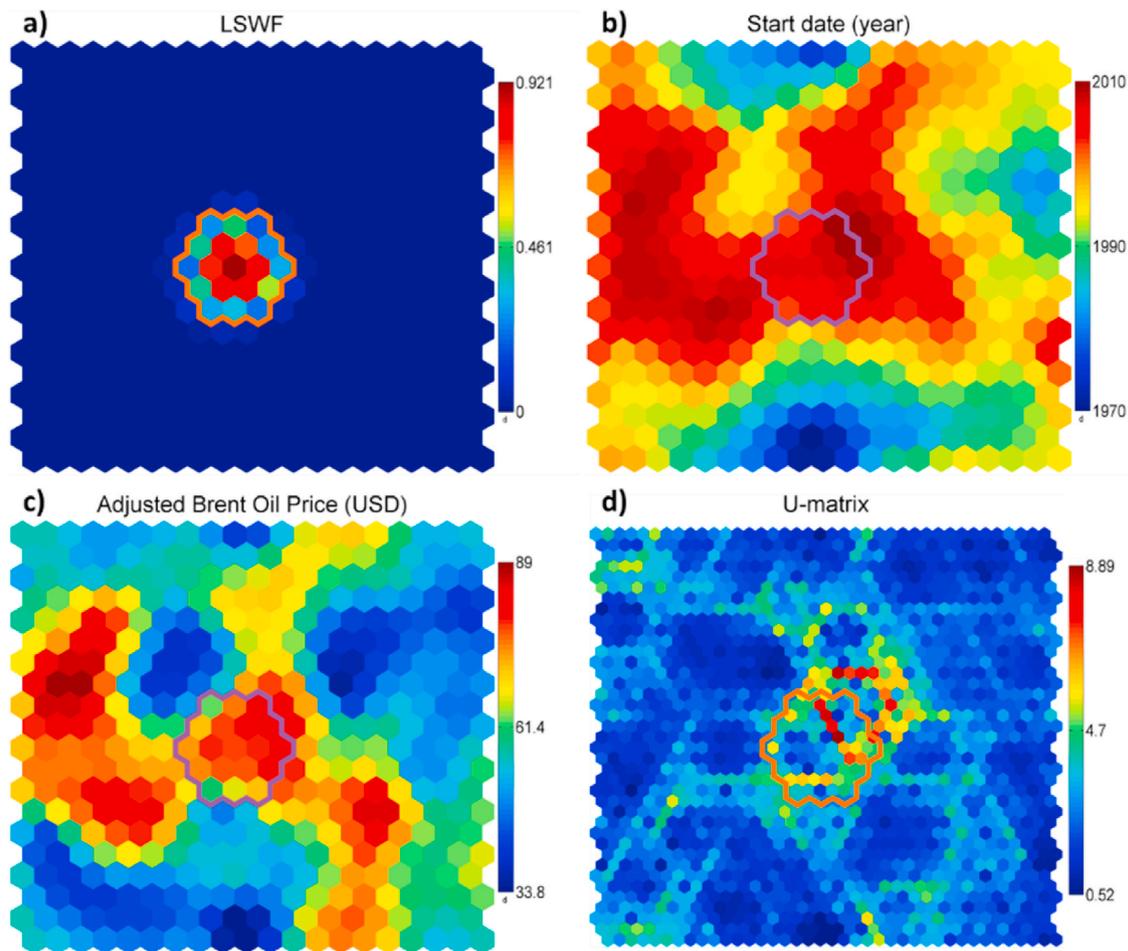


Fig. 38. Component plots related to a) the low salinity water injection method (LSWF); b) start date of the project; c) price of the barrel at the beginning of the project adjusted for inflation; d) U-matrix.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank National Agency for Petroleum, Natural Gas and Biofuels (ANP) and Petrobras S.A. for awarding a master's scholarship to the first author, the Integrated Technology for Rock and Fluid Analysis (InTRA), and the Graduate Program of the Department of Naval and Oceanic Engineering of the Polytechnic School of the University of São Paulo (PPGEN).

References

- Abidin, A.Z., Puspasari, T., Nugroho, W.A., 2012. Polymers for enhanced oil recovery technology. *Procedia Chem.* 4, 11–16. <https://doi.org/10.1016/j.proche.2012.06.002>.
- Aggarwal, C.C., 2015. *Data Mining*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-14142-8>.
- AlSofi, A.M., AlKhatib, A.M., Al-Ajwad, H.A., Wang, Q., Zahrani, B.H., 2019. Assessment of enhanced-oil-recovery-chemicals production and its potential effect on upstream facilities. *SPE J.* 24, 1037–1056. <https://doi.org/10.2118/183159-PA>.
- Alvarado, V., Manrique, E., 2010a. Enhanced oil recovery: an update review. *Energies* 3, 1529–1575. <https://doi.org/10.3390/en3091529>.
- Alvarado, V., Manrique, E., 2010b. *Enhanced Oil Recovery: Field Planning and Development Strategies*. Gulf Professional Publishing.
- Awan, A.R., 2008. *A Survey of North Sea Enhanced-Oil-Recovery Projects Initiated during the Years 1975 to 2005* 17.
- Babadagli, T., 2007. Development of mature oil fields — a review. *J. Petrol. Sci. Eng.* 57, 221–246. <https://doi.org/10.1016/j.petrol.2006.10.006>.
- Babadagli, T., 2020. Philosophy of EOR. *J. Petrol. Sci. Eng.* 188, 106930. <https://doi.org/10.1016/j.petrol.2020.106930>.
- Belazreg, L., Mahmood, S.M., 2020. Water alternating gas incremental recovery factor prediction and WAG pilot lessons learned. *J. Pet. Explor. Prod. Technol.* 10, 249–269. <https://doi.org/10.1007/s13202-019-0694-x>.
- Belhaj, A.F., Elraies, K.A., Mahmood, S.M., Zulkifli, N.N., Akbari, S., Hussien, O.S., 2020. The effect of surfactant concentration, salinity, temperature, and pH on surfactant adsorption for chemical enhanced oil recovery: a review. *J. Pet. Explor. Prod. Technol.* 10, 125–137. <https://doi.org/10.1007/s13202-019-0685-y>.
- Bertini, E., Tatu, A., Keim, D., 2011. Quality metrics in high-dimensional data visualization: an overview and systematization. *IEEE Trans. Vis. Comput. Graph.* 17, 2203–2212. <https://doi.org/10.1109/TVCG.2011.229>.
- Bondor, P.L., 2005. *Planning EOR Projects in Offshore Oil Fields*, vol. 7.
- Brodie, J., Zhang, P., Mellemstrand Hetland, S., Moulds, T., Jhaveri, B., 2012. BP North Sea gas injection projects: sustaining offshore production. Abu Dhabi, UAE. In: *All Days*. Presented at the Abu Dhabi International Petroleum Conference and Exhibition, SPE. <https://doi.org/10.2118/161189-MS>. SPE-161189-MS.
- Brodie, J.A., Jhaveri, B.S., Moulds, T.P., Mellemstrand Hetland, S., 2012. Review of gas injection projects in BP. *Society of Petroleum Engineers*. <https://doi.org/10.2118/154008-MS>.
- Chavan, M., Dandekar, A., Patil, S., Khataniar, S., 2019. Low-salinity-based enhanced oil recovery literature review and associated screening criteria. *Pet. Sci.* 16, 1344–1360. <https://doi.org/10.1007/s12182-019-0325-7>.
- Chen, Q., Otaibi, M., Ayirala, S., Yousef, A., 2021. The prospects and potential opportunities of low salinity water flooding for offshore applications in sandstones. *J. Petrol. Sci. Eng.* 199, 108260. <https://doi.org/10.1016/j.petrol.2020.108260>.
- Da Costa Fraga, C.T., Capeleiro Pinto, A.C., Branco, C.C.M., De Sant'Anna Pizarro, J.O., Da Silva Paulo, C.A., 2015. Brazilian pre-salt: an impressive journey from plans and challenges to concrete results. In: *All Days*. Presented at the Offshore Technology Conference, OTC, Houston, Texas, USA. <https://doi.org/10.4043/25710-MS>. OTC-25710-MS.
- De Moraes Cruz, R.O., Rosa, M.B., Branco, C.C.M., De Sant'Anna Pizarro, J.O., De Souza Silva, C.T., 2016. Lula NE pilot project - an ultra-deep success in the Brazilian pre-salt. In: *Day 1 Mon, May 02, 2016*. Presented at the Offshore Technology Conference, OTC, Houston, Texas, USA. <https://doi.org/10.4043/27297-MS>. D011S007R001.
- Delamaide, E., Degre, G., n.d. *Chemical EOR for Heavy Oil: the Canadian Experience* 31..

- Dong, X., Liu, H., Chen, Z., Wu, K., Lu, N., Zhang, Q., 2019. Enhanced oil recovery techniques for heavy oil and oilsands reservoirs after steam injection. *Appl. Energy* 239, 1190–1211. <https://doi.org/10.1016/j.apenergy.2019.01.244>.
- Fakher, S., Imqam, A., 2020. A data analysis of immiscible carbon dioxide injection applications for enhanced oil recovery based on an updated database. *SN Appl. Sci.* 2, 448. <https://doi.org/10.1007/s42452-020-2242-1>.
- Ganguli, S., Sompolinsky, H., 2012. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* 35, 485–508. <https://doi.org/10.1146/annurev-neuro-062111-150410>.
- Gbadamosi, A.O., Kiwalabye, J., Junin, R., Augustine, A., 2018a. A review of gas enhanced oil recovery schemes used in the North Sea. *J. Pet. Explor. Prod. Technol.* 8, 1373–1387. <https://doi.org/10.1007/s13202-018-0451-6>.
- Gbadamosi, A.O., Kiwalabye, J., Junin, R., Augustine, A., 2018b. A review of gas enhanced oil recovery schemes used in the North Sea. *J. Pet. Explor. Prod. Technol.* 8, 1373–1387. <https://doi.org/10.1007/s13202-018-0451-6>.
- Gbadamosi, A.O., Junin, R., Manan, M.A., Agi, A., Yusuff, A.S., 2019a. An overview of chemical enhanced oil recovery: recent advances and prospects. *Int. Nano Lett.* 9, 171–202. <https://doi.org/10.1007/s40089-019-0272-8>.
- Gbadamosi, A.O., Junin, R., Manan, M.A., Agi, A., Yusuff, A.S., 2019b. An overview of chemical enhanced oil recovery: recent advances and prospects. *Int. Nano Lett.* 9, 171–202. <https://doi.org/10.1007/s40089-019-0272-8>.
- Hashemi-Kiasari, H., Hemmati-Sarapardeh, A., Mighani, S., Mohammadi, A.H., Sedaee-Sola, B., 2014. Effect of operational parameters on SAGD performance in a dip heterogeneous fractured reservoir. *Fuel* 122, 82–93. <https://doi.org/10.1016/j.fuel.2013.12.057>.
- Huber, S., Wiemer, H., Schneider, D., Ihlenfeldt, S., 2019. DMME: data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP* 79, 403–408. <https://doi.org/10.1016/j.procir.2019.02.106>.
- Jerauld, G.R., Salter, S.J., 1990. The effect of pore-structure on hysteresis in relative permeability and capillary pressure: pore-level modeling. *Transp Porous Med* 5, 103–151. <https://doi.org/10.1007/BF00144600>.
- Ji-hong, H., Wen-long, G., Chang-feng, X.L., Hong-jie, C., Xiao-ling, L.L., Yu-xin, W., 2020. Production performance of in-situ combustion in heavy oil reservoir after steam injection. *Xinjiang Petroleum Geology* 31, 517.
- Jin, L., 2017. *Advancing CO2 Enhanced Oil Recovery and Storage in Unconventional Oil Play—Experimental Studies on Bakken Shales* 13.
- Kang, P.-S., Lim, J.-S., Huh, C., 2016. Screening criteria and considerations of offshore enhanced oil recovery. *Energies* 9, 44. <https://doi.org/10.3390/en9010044>.
- Karacan, C.Ö., Brennan, S.T., Buursink, M.L., Freeman, P.A., Lohr, C.D., Merrill, M.D., Olea, R.A., Warwick, P.D., 2023. A residual oil zone (ROZ) assessment methodology with application to the central basin platform (Permian Basin, USA) for enhanced oil recovery (EOR) and long-term geologic CO2 storage. *Geoenery Science and Engineering* 230, 212275. <https://doi.org/10.1016/j.geoen.2023.212275>.
- Ke, C.-Y., Lu, G.-M., Li, Y.-B., Sun, W.-J., Zhang, Q.-Z., Zhang, X.-L., 2018. A pilot study on large-scale microbial enhanced oil recovery (MEOR) in Baolige Oilfield. *Int. Biodeterior. Biodegrad.* 127, 247–253. <https://doi.org/10.1016/j.ibiod.2017.12.009>.
- Kohonen, T., 2001. Learning vector quantization. In: Kohonen, T. (Ed.), *Self-Organizing Maps*, Springer Series in Information Sciences. Springer, Berlin, Heidelberg, pp. 245–261. https://doi.org/10.1007/978-3-642-56927-2_6.
- Kohonen, T., 2013. Essentials of the self-organizing map. *Neural Network.* 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>.
- Kokal, S., Al-Kaabi, A., 2010. *Enhanced Oil Recovery: Challenges & Opportunities*. World Petroleum Council. Official Publication 2010 6.
- Kolodzie Jr., S., 1980 Sep. Analysis of pore throat size and use of the Waxman-Smits equation to determine OOIP in Spindle Field, Colorado. In: *SPE Annual Technical Conference and Exhibition? SPE pp. SPE-9382*.
- Kolodzie, S., 1980. Analysis of pore throat size and use of the waxman-smits equation to determine ooiip in spindle field, Colorado. *OnePetro*. <https://doi.org/10.2118/9382-MS>.
- Kowalewski, E., Rueslåtten, I., Steen, K.H., Bødtger, G., Torsæter, O., 2006. Microbial improved oil recovery—bacterial induced wettability and interfacial tension effects on oil production. *Journal of Petroleum Science and Engineering, Reservoir Wettability* 52, 275–286. <https://doi.org/10.1016/j.petrol.2006.03.011>.
- Kumar, V.K., Fassihi, M.R., 1995. Case history and appraisal of the medicine Pole hills unit air injection project. *SPE Reservoir Eng.* 10, 198–202. <https://doi.org/10.2118/27792-PA>.
- Lake, L., Lotfollahi, M., Bryant, S., 2019. Fifty years of field observations: lessons for Co2 storage from Co2 enhanced oil recovery. *SSRN Journal*. <https://doi.org/10.2139/ssrn.3366254>.
- Lamas, L.F., Botechia, V.E., Correia, M.G., Schiozer, D.J., Delshad, M., 2018. Influence of polymer properties on selection of production strategy for a heavy oil field. *J. Petrol. Sci. Eng.* 163, 110–118. <https://doi.org/10.1016/j.petrol.2017.12.074>.
- Le, J.-J., Wu, X.-L., Wang, R., Zhang, J.-Y., Bai, L.-L., Hou, Z.-W., 2015. Progress in pilot testing of microbial-enhanced oil recovery in the Daqing oilfield of north China. *Int. Biodeterior. Biodegrad.* 97, 188–194. <https://doi.org/10.1016/j.ibiod.2014.10.014>.
- Li, P., Yi, L., Liu, X., Hu, G., Lu, J., Zhou, D., Hovorka, S., Liang, X., 2019. Screening and simulation of offshore CO2-EOR and storage: a case study for T the HZ21-1 oilfield in the pearl river mouth basin, northern South China Sea. *Int. J. Greenh. Gas Control* 16.
- Liu, Z., Liang, Y., Wang, Q., Guo, Y., Gao, M., Wang, Z., Liu, W., 2020. Status and progress of worldwide EOR field applications. *J. Petrol. Sci. Eng.* 193, 107449. <https://doi.org/10.1016/j.petrol.2020.107449>.
- Manrique, E., 2009. *Effective EOR Decision Strategies with Limited Data: Field Cases Demonstration* 11.
- Manrique, E.J., Kamouei, M.I., Kitchen, C.D., Alvarado, V., 2008. Effective EOR decision strategies with limited data: field cases demonstration. In: Presented at the SPE Symposium on Improved Oil Recovery. *OnePetro*. <https://doi.org/10.2118/113269-MS>.
- Miller, R.J., 1995. *Koch's Experience with Deep in Situ Combustion in Williston Basin (No. NIPER/BDM-0086; CONF-940450-)*. BDM Oklahoma, Inc., Bartlesville, OK (United States). National Inst. for Petroleum and Energy Research, Bartlesville, OK (United States).
- Mokheimer, E.M.A., Hamdy, M., Abubakar, Z., Shakeel, M.R., Habib, M.A., Mahmoud, M., 2019. A comprehensive review of thermal enhanced oil recovery: techniques evaluation. *J. Energy Resour. Technol.* 141, 030801. <https://doi.org/10.1115/1.4041096>.
- Morel, D.C., Zaugg, E., Jouenne, S., Danquigny, J.A., Cordelier, P.R., 2015. Dalia/camelia polymer injection in deep offshore field Angola learnings and in situ polymer sampling results. In: Day 3 Thu, August 13, 2015. Presented at the SPE Asia Pacific Enhanced Oil Recovery Conference. SPE, Kuala Lumpur, Malaysia. <https://doi.org/10.2118/174699-MS>. D031S014R001.
- Muggeridge, A., Cockin, A., Webb, K., Frampton, H., Collins, I., Moulds, T., Salino, P., 2014. Recovery rates, enhanced oil recovery and technological limits. *Proc. R. Soc. A* 372, 20120320. <https://doi.org/10.1098/rsta.2012.0320>.
- Murer, A.S., McClennen, K.L., Ellison, T.K., Timmer, R.S., Larson, D.C., Wolcott, K.D., Walker, T.M., Thomsen, M.A., 1997. Steam Injection Project in Heavy Oil Diatomite. *OnePetro*. <https://doi.org/10.2118/38302-MS>.
- Olajire, A.A., 2014. Review of ASP EOR (alkaline surfactant polymer enhanced oil recovery) technology in the petroleum industry: prospects and challenges. *Energy* 77, 963–982. <https://doi.org/10.1016/j.energy.2014.09.005>.
- Patel, J., Borgohain, S., Kumar, M., Rangarajan, V., Somasundaran, P., Sen, R., 2015. Recent developments in microbial enhanced oil recovery. *Renew. Sustain. Energy Rev.* 52, 1539–1558. <https://doi.org/10.1016/j.rser.2015.07.135>.
- Piatetsky, Gregory, 2014. *CRISP-DM, Still the Top Methodology for Analytics, Data Mining, or Data Science Projects*. KDD News [WWW Document].
- Raney, K., Ayirala, S., Chin, R., Verbeek, P., 2012. Surface and subsurface requirements for successful implementation of offshore chemical enhanced oil recovery. *SPE Prod. Oper.* 27, 294–305. <https://doi.org/10.2118/155116-PA>.
- Rassenfoss, S., 2011. From bacteria to barrels: microbiology having an impact on oil fields. *J. Petrol. Technol.* 63, 32–38. <https://doi.org/10.2118/1111-0032-JPT>.
- Reksidler, R., Vieira, R.A.M., Orlando, A.E., Costa, B.R.S., Pereira, L.S., 2015. Offshore chemical enhanced oil recovery. In: *OTC Brasil. Offshore Technology Conference*. <https://doi.org/10.4043/26123-MS>. Rio de Janeiro, Brazil.
- Rezaee, M.R., Jafari, A., Kazemzadeh, E., 2006. Relationships between permeability, porosity and pore throat size in carbonate rocks using regression analysis and neural networks. *J. Geophys. Eng.* 3, 370–376. <https://doi.org/10.1088/1742-2132/3/4/008>.
- Shafique, U., Qaiser, H., 2014. A comparative study of data mining process models (KDD, CRISP-DM and SEMMA) 12, 7.
- Sheng, J., 2010. *Modern Chemical Enhanced Oil Recovery: Theory and Practice*. Gulf Professional Publishing.
- Taber, J.J., Martin, F.D., Seright, R.S., 1997. EOR screening criteria revisited - Part 1: introduction to screening criteria and enhanced recovery field projects. *SPE Reservoir Eng.* 12, 189–198. <https://doi.org/10.2118/35385-PA>.
- Thomas, S., 2008. Enhanced oil recovery - an overview. *Oil & Gas Science and Technology - Rev. IFP* 63, 9–19. <https://doi.org/10.2516/ogst:2007060>.
- Turta, A.T., Chattopadhyay, S.K., Bhattacharya, R.N., Condrachi, A., Hanson, W., 2007. Current status of commercial in situ combustion projects worldwide. *J. Can. Petrol. Technol.* 46, 7.
- Vishnyakov, V., Suleimanov, B., Salmanov, A., Zeynalov, E., 2020. *Primer on Enhanced Oil Recovery*. Gulf Professional Publishing, Cambridge, MA.
- Wang, L., Fu, X., 2018. Data-driven analyses of low salinity water flooding in sandstones. *Fuel* 234, 674–686. <https://doi.org/10.1016/j.fuel.2018.07.063>.
- Warwick, P.D., Attanasi, E., Blondes, M., Brennan, S., Buursink, M., Doolan, C., Freeman, P., Jahediesfanjani, H., Karacan, C.Ö., Lohr, C., Merrill, M., Olea, R., Roueché, J., Shelton, J., Slucher, E., Varela, B.A., Verma, M.K., 2019. Carbon dioxide enhanced oil recovery and residual oil zone studies at the U.S. Geological survey. *SSRN Journal*. <https://doi.org/10.2139/ssrn.3366202>.
- Wever, D.A.Z., Picchioni, F., Broekhuis, A.A., 2011. Polymers for enhanced oil recovery: a paradigm for structure–property relationship in aqueous solution. *Prog. Polym. Sci.* 36, 1558–1628. <https://doi.org/10.1016/j.proprogmsci.2011.05.006>.
- Wirth, R., Hipp, J., 2000. *CRISP-DM: Towards a Standard Process Model for Data Mining* 11.
- Yuan, S., Wang, Q., 2018. New progress and prospect of oilfields development technologies in China. *Petrol. Explor. Dev.* 45, 698–711. [https://doi.org/10.1016/S1876-3804\(18\)30073-9](https://doi.org/10.1016/S1876-3804(18)30073-9).
- Yuan, B., Wood, D.A., 2018. A comprehensive review of formation damage during enhanced oil recovery. *J. Petrol. Sci. Eng.* 167, 287–299. <https://doi.org/10.1016/j.petrol.2018.04.018>.
- Zhang, N., Wei, M., Bai, B., 2018. Statistical and analytical review of worldwide CO2 immiscible field applications. *Fuel* 220, 89–100. <https://doi.org/10.1016/j.fuel.2018.01.140>.

Anexo C – Artigo publicado na revista *Energy and AI*



Stepped machine learning for the development of mineral models: Concepts and applications in the pre-salt reservoir carbonate rocks

Lucas Abreu Blanes de Oliveira^{a,b,*}, Luiz Felipe Niedermaier Custódio^b,
Thais Bortotti Fagundes^b, Carina Ulsen^b, Cleyton de Carvalho Carneiro^b

^a Petrobras – Petróleo Brasileiro S.A., Avenida Henrique Valadares 28, Centro, Rio de Janeiro, 20231-030 Rio de Janeiro, Brazil

^b Universidade de São Paulo, USP, Escola Politécnica, Praça Narciso de Andrade, Vila Mathias, Santos, São Paulo 11013-560, Brazil

HIGHLIGHTS

- Use of machine learning for the development of mineralogical models.
- Innovative stepped training to estimate mineral concentrations.
- Reliable mineralogical model to be used in formation evaluation.

ARTICLE INFO

Article history:

Received 2 December 2020

Received in revised form 18 January 2021

Accepted 20 January 2021

Available online 26 January 2021

ABSTRACT

Understanding rock mineralogy is essential for formation evaluation, improving the calculation of porosity and hydrocarbon saturation. The primary method to obtain the mineralogy from a well is by applying a model to the geochemical tool's chemical elements. However, creating a mineralogical model presents challenges such as the minerals' chemical composition and the decision to include a mineral in the model. The traditional application of machine learning can make mineral models less realistic since conventional training is developed based on a set of minerals with different occurrences, lowering some minerals' representativeness. The present research proposes the stepped machine learning (SML), a stepped way to use machine learning to create a mineralogical model from chemical and mineralogical data. A database was assembled with the elemental concentration obtained with XRF analyses and the mineral concentrations obtained with XRD analyses. The chemical elements were Al, Ca, Fe, K, Mg, Mn, Na, Si, and Ti. The minerals were calcite, dolomite, quartz, clays, K-feldspar, plagioclase, and pyroxene. Four algorithms were tested: MLP, GAN, Random Forest, and XGBoost, with XGBoost showing the best results. SML was applied, where a mineral model results are used to train a subsequent model. SML allowed for a significant improvement in some models, notably to clays with an increase in R^2 from 0.597 to 0.853, quartz an increase from 0.673 to 0.869, and calcite, from 0.758 to 0.862. A decrease in the mean squared error of these minerals' models was also observed. The model was applied to the geochemical logs from three wells drilled in the Brazilian pre-salt, and the results were compared with XRD analyzes. The SML model was able to honor the mineral concentrations for different rocks. It is demonstrated that the integration between machine learning tools and geological knowledge in SML was crucial for creating a representative mineralogical model.

Introduction

Understanding the mineralogy of the rock matrix is essential for formation evaluation. A reliable mineralogical model can significantly improve the calculation of porosity, hydrocarbon saturation, and volume of clay [10], also providing information on the cation exchange capacity [17] and total organic carbon [13]. In addition to the direct application in formation evaluation, knowledge of mineralogy can be used to assist acid-fracturing operations [21] and to monitor the variation of oil/water contact during the production of a hydrocarbon reservoir [24,28,30].

The primary way to obtain a mineralogical model of a reservoir is by using the geochemical data acquired in well logging operations. The

geochemical tool detects chemical elements present in the rock matrix, exploring the spectrum of gamma-rays generated by the interaction between neutrons emitted by the tool and these elements [8]. Concentrations of Al, Ca, Fe, K, Mg, Na, and Si, can be obtained and represents the main minerals present in reservoir rocks. However, as it is an indirect measure, the concentrations of chemical elements and their respective allocations in different minerals must be marked out with information obtained from the reservoir's rock samples.

According to Herron et al. [18], elemental concentrations of rock samples can be measured by X-ray fluorescence (XRF) or by inductively coupled plasma atomic emission spectrometry (ICP-AES), while mineralogy can be analyzed by X-ray diffraction (XRD). The construction

* Corresponding author at: Petrobras – Petróleo Brasileiro S.A., Avenida Henrique Valadares 28, Centro, Rio de Janeiro, 20231-030 Rio de Janeiro, Brazil.
E-mail address: lucas.oliveira@usp.br (L.A.B. de Oliveira).



Fig. 1. Location of the Brazilian pre-salt and the area of study.

of a mineralogical model would then consist of a mathematical inversion that would transform chemical elements' concentrations into mineralogical concentrations [2,9,32]. However, creating this mathematical model is not a simple task; as pointed out by Freedman et al. [10], the minerals chemical composition, their respective endpoints, and the subjective assumption of a specific mineral's presence or absence (such as clay or trace minerals) must serve as input for the model.

In light of the difficulties presented, this research seeks to develop a mineralogical model for the carbonate, siliciclastic, and igneous rocks of the Brazilian pre-salt using machine learning algorithms. The main advantage of machine learning is that it does not need prior knowledge of the minerals' composition, reducing the model's subjectivity. The chemical concentrations of rock samples obtained through XRF measurements will be used as inputs for the model, and the mineralogy obtained in XRD will be the outputs. The model will include the main minerals observed in the pre-salt formations. Additionally, an innovative training sequence proposed in this research, called stepped machine learning (SML), will improve the models' predictive capacity. In the end, the model generated by SML will be used in the mineralogical interpretation of a well drilled in the pre-salt.

The present work is organized as follows:

- **Conceptual background:** the main aspects of the Brazilian pre-salt rocks are discussed, followed by a review of the XRF and XRD analyses and the machine learning algorithms used.
- **Methodology:** the workflow is presented from the XRF and XRD analyzes, the construction of the database, and the process of training and evaluation of the machine learning models.
- **Results:** the results of the trained models are shown, as well as the importance of each of the input variables during training. Then, the application of the models to estimate the mineralogy of rocks in a well is presented.
- **Discussion and conclusion:** the final configuration of the SML models and the main conclusions are summarized.

Conceptual background

Pre-salt rocks

Located in Santos Basin (Fig. 1), the Brazilian pre-salt consists of the rocks found stratigraphically below the thick layer of evaporites of the Ariri Formation of the Neo-Aptian age [23]. This sequence includes the Barra Velha Formation's stromatolites, the grainstones, wackestones,

and packstones from the Itapema Formation, sandstones, pelites shales from the Piçarras Formation, and the igneous rocks of the Camboriú Formation. Intercalations with basalts can also occur. The sedimentary rocks were deposited between the Hauterivian and the Eo-Aptian (between approximately 110–130 million years), while the crystalline basement has Precambrian age.

XRF and XRD analyses

The X-ray fluorescence (XRF) is characterized by the emission of secondary X-rays after a particular material is bombarded by high-energy X (primary emission). XRF is widely used to analyze materials' chemical composition since the radiation energy emitted will depend on the atoms/electron transition from a given material. When exposed to high-energy electromagnetic radiation in short-wavelength X-rays, atoms may be ionized. During ionization, electrons present in the inner orbits of atoms are ejected, making them unstable and causing electrons from outer orbits to replace them. The change between orbits causes photons' emission, whose energy equals the energy difference between the two orbits, an intrinsic characteristic of the atom involved. The re-emission of radiation in energy other than that initially absorbed is called fluorescence [27]. The radiation emitted by the material after exposure to X-rays is acquired in a spectrum, where energy peaks represent different chemical elements. The intensity of these peaks allows the quantification of the elements that compose the measured material.

X-ray diffraction (XRD) corresponds to one of the main techniques for crystalline materials' microstructural characterization, responsible for providing qualitative and quantitative results of the crystalline phases present [6]. It represents the phenomenon of interaction between the incident X-ray beam and the electrons of the atoms that make up a material, related to coherent scattering. If the atoms that generate this scattering are arranged systematically, as in a crystalline structure, the phase relationships between the scattering become periodic, and the X-ray diffraction phenomenon is observed at various angles of incidence. The various planes of a crystalline structure have different atoms and electrons' densities, making the diffracted intensities specific to their different planes. Thus, each crystalline compound has a unique diffractometric pattern, and its correct identification occurs through the angular positions and the relative intensities of the diffracted beams. However, this will only occur when the incident X-ray satisfies Bragg's Law [20]. The diffractometric pattern, through angular positions and relative intensities, is compared with reference standards, available in updated databases and made available by the ICDD - International Center for Diffraction Data and ICSD - Inorganic Crystal Structure Database, obtaining crystallographic information and physical properties of the crystalline compounds (qualitative analysis).

Machine learning algorithms

Several machine learning algorithms can be found in the literature, using the most different training strategies. However, some algorithms stand out for their intense application in recent years. Among them, it is possible to mention artificial neural networks (ANN) and tree-based algorithms.

ANN can be considered the most popular machine learning algorithm today, vastly applied in deep learning problems. It uses artificial neurons at its core. Neurons receive input data and calculate outputs using activation functions capable of capturing non-linearity. The most common ANN structure is Multilayer Perceptron (MLP, [11,26,29]). It is a feedforward arrangement containing layers of multiple neurons. The training process updates the weights assigned to each neuron connection, using back-propagation to converge to the minimum error.

Generative Adversarial Networks (GAN, [14]) are a class of ANN where two networks compete against each other in a zero-sum game. A discriminator receives input data from two sources: the actual data and data created by a generator. At each iteration, the discriminator's

function is to identify which information is real and synthetic created by the generator. The function of the generator is to provide information to the discriminator to deceive it. Initially, the discriminator can easily distinguish the synthetic data. However, throughout training, both the generator and the discriminator evolve to represent the actual data. In the end, the discriminator finds it very difficult to identify the real data of the synthetic provided by the generator, which has learned the patterns of the input data and can imitate them perfectly. The discriminator and generator structures can vary, from an MLP to a convolutional or recurrent network. GANs have been applied to deep learning problems, mainly involving images. Aggarwal et al. [1] proposed a GAN architecture for regression problems, comparing its performance with other algorithms. They identified that GANs performed better on data with heteroscedasticity. In data with homoscedasticity, tree-based algorithms showed better results.

Random Forest is an ensemble algorithm that uses bagging (bootstrap aggregation) and several decision trees [3,4,19]. The technique known as bagging divides the data into subsets and fits several decision trees, combining their predictions. This combination reduces the variance that would be observed in a single tree. To further reduce variance, Random Forest randomly selects each subset's input variables, making the model less susceptible to overfitting.

Gradient Boosting uses the union of several simple predictors, also called weak learners [12]. These weak learners are trained in sequence so that a predictor learns using the previous one's error. In this way, the sequence of weak learners forms a robust committee [15]. In Gradient Boosting, the weak learner used is the decision tree. A tree fits the residuals from the previous iteration during training, converging to the smallest possible error. This sequence can be understood as a gradient descent method and can be summarized as follows [22]:

Considering y as the instances, $F(x)$ as the predicted value of y , and $h(x)$ as the decision tree.

- Calculate the first approximation for $F_0(x)$, such as the average of y .
- For i iterations, do:
 - Calculate the residuals $r_i = y - F_{i-1}(x)$.
 - Fit a tree $h_i(x)$ to the residuals.
 - Add the tree to the model, as $F_i(x) = F_{i-1}(x) + h_i(x)$.
- Repeat the process until the error becomes constant or until the end of iterations.

An advantage of Gradient Boosting is the ability to calculate the importance of variables during training. It can be calculated by counting the number of times a variable is used to split a node in a decision tree. This importance can be used to provide insight into how each variable was used in the training stage.

Several metrics can be used to assess the quality of machine learning models. In regression problems, most common are the coefficient of determination (R^2) and the mean squared error (MSE). The R^2 expresses how much of the dependent variable's variance can be predicted by the independent variable and can vary between 0 and 1. In machine learning, this comparison is made between real and modeled data, and a R^2 close to 1 indicates that the algorithm can reproduce the real data with high confidence.

The mathematical definition of R^2 is given by Eqs. 1, 2, and 3:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (1)$$

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad (2)$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (3)$$

Where SS_{res} is the total sum of squares of residuals, expressed by the square of the difference between the modeled (f_i) and real (y_i) values; and SS_{tot} is the total sum of squares, expressed by the square of the difference between the real data and its mean (\bar{y}). If the R^2 between the

Table 1

Summary of the samples collected in the pre-salt formations.

Formation	Number of samples	Lithotypes	Percentage
Barra Velha	1059	Carbonate, igneous	77.0%
Itapema	288	Carbonate, siliciclastic, igneous	20.8%
Piçarras	9	Siliciclastic, igneous	0.7%
Camboriú	20	Igneous	1.5%

real and modeled data is 1, the model perfectly predicts the real data. If R^2 is 0, the model prediction is as good as the mean of the data. Negative R^2 indicates that the model is worse than the mean.

The MSE indicates the average error of the model, calculated by Equation 4:

$$MSE = \frac{1}{n} \sum_i (y_i - f_i)^2 \quad (4)$$

Where n is the total number of instances. The MSE will always be positive, and values close to 0 indicate a high-quality model. Unlike R^2 , the MSE does not have a maximum value.

Methodology

The database consists of analysis from 1,376 instances of rock samples collected from 22 wells drilled in the Brazilian pre-salt. From this total, 77 samples are cuttings collected during well drilling, 199 are rotary sidewall cores collected during wireline operations, and 1,100 are core plugs. Table 1 shows the number of samples collected in each of the pre-salt formations and the lithotypes observed in each formation.

First, the samples were sent to the XRF and XRD laboratories, where chemical and mineralogical composition measurements were carried out. The results were concatenated in a database, where the chemical composition served as input, and the mineralogy served as output for the training process. Four algorithms were tested and compared to choose the best mineral model. The training and evaluation of the model followed the best practices for creating a machine learning model. The complete workflow is shown in Fig. 2 and is detailed below.

XRF and XRD laboratory analyses

The XRF analyzes were performed at the Technological Characterization Laboratory (LCT) of the University of São Paulo and by the company SGS Geosol. For cuttings and core plugs, the chemical determination was carried out by the quantitative method, using specific calibration curves for carbonate rocks, with the quantification of CaO, MgO, SiO₂, Al₂O₃, Fe₂O₃, Na₂O, K₂O, P₂O₅, TiO₂, SrO, MnO and S. The determination of loss on ignition (LOI) was carried out at 1,020 °C for two hours. The analyzes were performed on the Zetium equipment from the PANalytical company. The preparation and analysis of the samples consisted of fused beads with the addition of lithium tetraborate.

For XRD analysis, the sample preparation comprised the pulverization of 20 g of material to below 0.04 mm, using a planetary ball mill, and manual backloading pressing in proper sample holders. The diffractograms collection was carried out in a Panalytical X'Pert equipment with a detector sensitive to the X'Celerator position with copper tube (Cu K α radiation - $\lambda = 1.54186 \text{ \AA}$ with Ni filter) at 45 kV \times 40 mA, 2θ ranging from 2 to 70°, 0.02° 2θ step size, time per step of 0.20 s, and total collection time of 53 s.

Database preparation

Chemical (XRF) and mineralogical (XRD) data were joined in a database, where the chemical elements were considered input variables, and the mineralogical concentrations were output variables. Table 2 summarizes the main characteristics of the variables, as well as their statistics.

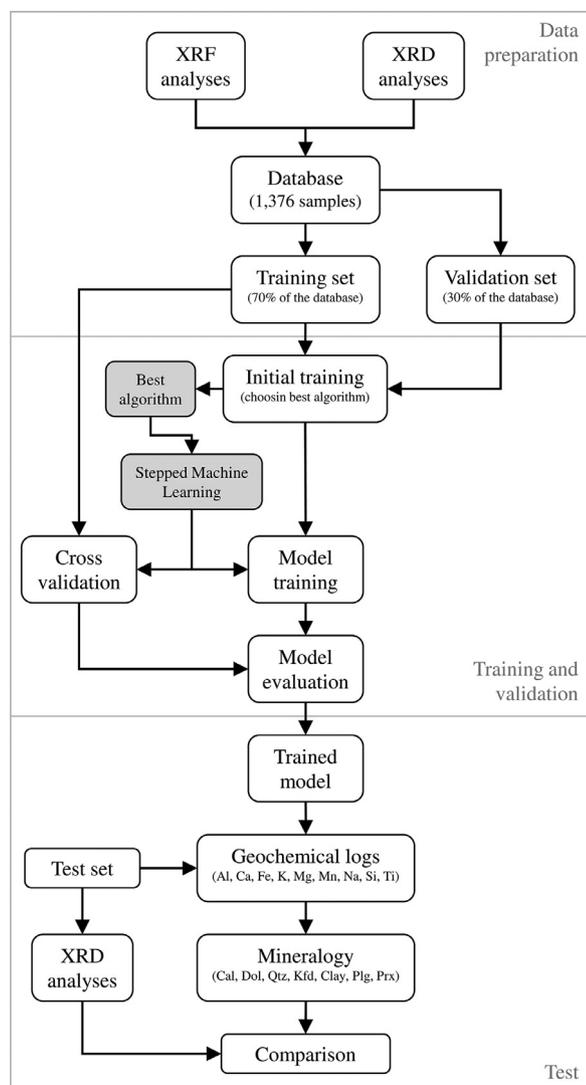


Fig. 2. Workflow used in the present research.

The elements selected as input were those acquired in wireline operations. The minerals selected as output were the most abundant in the pre-salt rocks, based on their maximum, average, and non-zero concentrations. Therefore, the minerals selected as output were calcite, dolomite, quartz, K-feldspar, clays, plagioclase, and pyroxene. These minerals are the main components of the carbonate, siliciclastic, and igneous rocks observed in the pre-salt. Minerals such as dawsonite, pyrite, barite, fluorite, hematite, and ilmenite are also present in low concentrations and were not included as output variables.

The chemical formula of these minerals, shown in Table 2, demonstrates how challenging it is to build a mineralogical model without machine learning. Minerals such as dolomite, K-feldspar, plagioclase, and pyroxene have complex compositions, with variations in the proportions of Ca–Mg, K–Al, Na–Ca, and Si–Al, making it challenging to choose endpoints. Despite having a more stable composition, calcite and quartz are subject to minor contamination by Fe, Mg, and Mn. Clays have complex compositions, which can vary according to the source rock and the environment in which the clay was deposited, making the choice of a single chemical formula impossible.

Fig. 3 presents the correlation between variables. It is possible to observe the high positive correlation between Ca and LOI, resulting from the burning of carbon dioxide from the carbonates, a phenomenon observed and described by studies such as that of Dean [7] and Heiri et al. [16]. The positive correlation between Al, Fe, K, and Ti reflects the chemical composition of siliciclastic and igneous rocks present in the pre-salt. The significant negative correlation between Ca and Si marks the passage of carbonate rocks from the Barra Velha and Itapema Formations to the siliciclastic and igneous rocks from the Piçarras and Camboriú Formations.

Concerning minerals, it is possible to observe the negative correlation between calcite and dolomite, a reflection of diagenetic dolomitization processes common in carbonates. The high positive correlation in plagioclase and pyroxene also stands out, indicating that these minerals probably compose the same rock types. Some obvious correlations are observed, such as Ca and calcite, Mg and dolomite, Si and quartz, and K and K-feldspar. The high positive correlation between Fe and Ti with plagioclase and pyroxene was not expected since these elements are not observed in these minerals' chemical composition. The clays did not significantly correlate with any chemical element, reinforcing the challenge of obtaining models that take them into account.

Table 2
Statistics of the input and output variables used in the training steps.

Input variables								
Name	Description	Unit	Min	Max	Mean	Std	Non-zero	
Al	Aluminum dry weight	g/g	0.000	0.088	0.006	0.013	99.3%	
Ca	Calcium dry weight	g/g	0.003	0.402	0.247	0.078	100.0%	
Fe	Iron dry weight	g/g	0.000	0.099	0.006	0.013	100.0%	
K	Potassium dry weight	g/g	0.000	0.090	0.004	0.009	97.1%	
Mg	Magnesium dry weight	g/g	0.000	0.179	0.045	0.026	100.0%	
Mn	Manganese dry weight	g/g	0.000	0.003	0.000	0.000	97.4%	
Na	Sodium dry weight	g/g	0.000	0.128	0.003	0.006	99.3%	
Si	Silicon dry weight	g/g	0.000	0.421	0.089	0.071	100.0%	
Ti	Titanium dry weight	g/g	0.000	0.021	0.001	0.002	95.6%	
LOI	Loss on ignition	g/g	0.008	0.479	0.349	0.094	100.0%	
Output variables								
Name	Description	Formula	Unit	Min	Max	Mean	Std	Non-zero
Cal	Calcite	CaCO ₃	g/g	0.00	1.00	0.48	0.25	93.8%
Dol	Dolomite	CaMg(CO ₃) ₂	g/g	0.00	1.00	0.29	0.21	94.7%
Qtz	Quartz	SiO ₂	g/g	0.00	0.98	0.16	0.14	93.4%
Kfd	K-Feldspar	KAlSi ₃ O ₈	g/g	0.00	0.64	0.02	0.05	26.7%
Clay	Clay	Variable	g/g	0.00	0.48	0.03	0.07	25.1%
Plg	Plagioclase	NaCa(Si ₃ AlO ₈) ₂	g/g	0.00	0.70	0.01	0.07	5.1%
Prx	Pyroxene	CaMgSi ₂ O ₆	g/g	0.00	0.34	0.00	0.03	3.3%
Others	Other minerals	Variable	g/g	0.00	0.31	0.01	0.02	23.5%

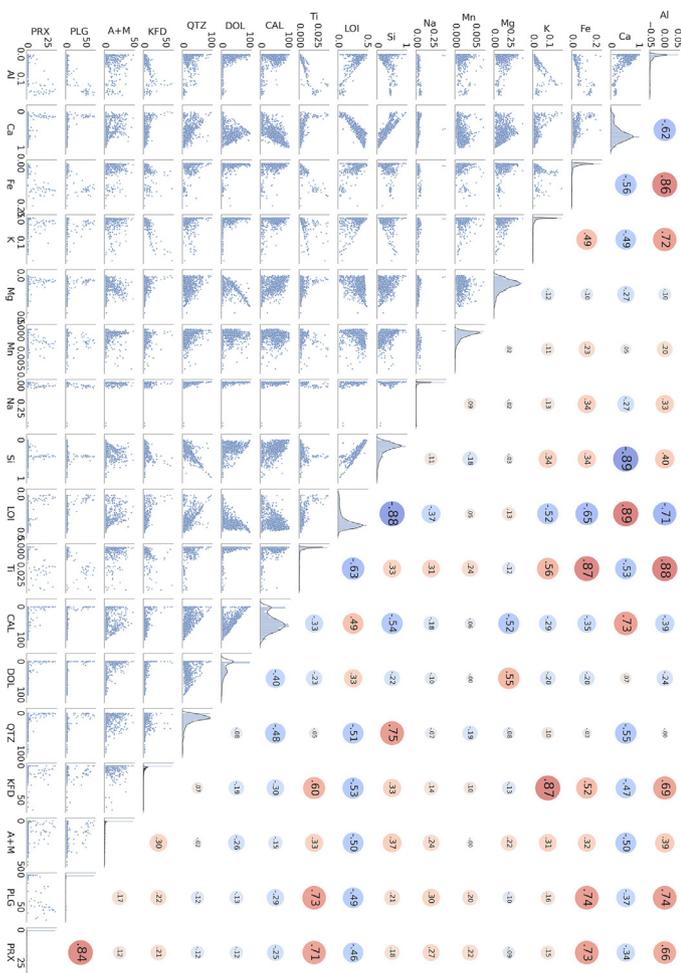


Fig. 3. Correlation between GAN variables, presented in the form of graphs and R^2 . Warm colors indicate a positive correlation, and cold colors indicate a negative correlation. Histograms of the variables are diagonally in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Model training

The database was randomly divided into a training set and a validation set. From the total data, 70% were separated for training (963 instances), and 30% were for validation (413 instances). As the concentrations of minerals and chemical elements vary between zero and one, data normalization/standardization was unnecessary. In addition to the minerals presented in Table 2, models for LOI, carbonates (calcite + dolomite), and pyroxene + plagioclase were also trained. LOI has a high correlation with carbonates and organic matter and is not acquired in wireline operations. The carbonate fraction of the rock, represented by calcite + dolomite, can be more easily fitted by the model since it correlates to LOI. As plagioclase and pyroxene have low abundance in the database and their concentrations have a high positive correlation, the combination of these minerals can present better results in training.

Before definitively creating the models, four machine learning algorithms were tested to assess the most suitable application in the database: Multilayer Perceptron (MLP), Generative Adversarial Networks (GAN), Random Forest, and Gradient Boosting. This choice aimed to cover both artificial neural networks and tree-based algorithms. The MLP used had the input layer, a hidden layer with 100 neurons, and the output layer. The GAN's structure was the one proposed by Aggarwal et al. [1], created to deal specifically with regression problems. Random

Forest had 100 decision trees without depth restriction or limiting the number of instances for each leaf and each split. The Gradient Boosting algorithm used was the one proposed by Chen & Guestrin [5], known as XGBoost.

As the first test's objective was to evaluate each of the algorithms' initial performance and then choose the best one to be used in the research, no hyperparameter tuning was performed. The algorithms were trained using the training set, generating models for the LOI and minerals. The trained algorithms estimated mineralogical concentrations using the input data from the validation set, and these concentrations were compared with the actual data using the R^2 metric. These results are shown in Table 3.

It is possible to observe that XGBoost presented the best results for most minerals, reflecting the highest mean R^2 . This result is in line with those obtained by Oliveira & Carneiro [25], who concluded that tree-based boosting algorithms present better results when applied to well data than MLP and Random Forest. GAN had the worst results, also in agreement with the results of Aggarwal et al. [1], which observed that GANs did not perform as well as algorithms like XGBoost in regression problems using tabular data. With these results, it was decided to use the XGBoost algorithm for the creation of mineralogical models. A fine-tuning found XGBoost's best hyperparameters, presented in Table 4. A detailed description of each hyperparameter can be found in XGBoost documentation [31].

The models were trained using the training set and then applied to the validation set. The results were evaluated using R^2 . In addition to the application in the validation set, 5-fold cross-validation was applied to the training sets. In this cross-validation, the training set was divided into five parts, where four were used to fit the model and one as validation. The process was repeated five times until all folds have been used for validation. The validation set and cross-validation allow an unbiased evaluation of the machine learning models' quality and robustness and reduce overfitting.

Stepped machine learning

As the mineralogical composition of a rock is a closed system, where the sum of the minerals is equal to one, it is expected that prior knowledge of the concentration of one mineral will impact the estimation of the others. Thus, the present research proposes the use of an innovative learning technique called stepped machine learning (SML) to estimate the mineralogical composition of rocks.

SML consists of sequential training of mineral models, adding one mineral's concentrations to the input variables during the others' training. However, it is not desirable to include all the previous results for training a given mineral model since it can carry the previous models' errors. Therefore, it is necessary to use the integration of geological and machine learning knowledge to choose which minerals will serve as input for subsequent models.

The step-by-step of SML was as follows:

LOI model: since LOI is not acquired by any wireline tool, a model was first created using chemical element concentrations. In a real well situation, these concentrations would come from the geochemical tool. Estimates of the other minerals were made using only the chemical elements for comparison with subsequent models.

Carbonate model: LOI was added to the inputs, and a model for carbonates (calcite + dolomite) was created. As LOI has a high correlation with carbonates, it is expected that this model has better results than the previous ones. Again, estimation of the other minerals was made using the chemical elements + LOI to be compared with later steps.

Mineral model: a new input was created by concatenating chemical elements + LOI + carbonates to create the other minerals' models. As carbonates represent the most important fraction of the pre-salt mineralogy, their inclusion in the model inputs can significantly improve other minerals' estimation.

Table 3
Results of the initial training phase of the machine learning algorithms.

Algorithm	LOI	Carb	Cal	Dol	Qtz	Kfd	Clay	Plg	Prx	Plg + Prx	Mean
MLP	0.836	0.817	0.711	0.682	0.663	0.703	0.409	0.954	0.614	0.864	0.725
GAN	0.816	0.625	0.308	0.457	0.313	0.143	0.225	0.187	0.285	0.317	0.367
Random Forest	0.811	0.759	0.695	0.668	0.652	0.650	0.528	0.918	0.760	0.765	0.721
XGBoost	0.895	0.811	0.774	0.791	0.656	0.802	0.602	0.959	0.641	0.833	0.776

Table 4
Best XGBoost hyperparameters for each model trained.

Hyperparameter name	LOI	Carb	Cal	Dol	Qtz	Kfd	Clay	Plg	Prx	Prx+Plg
Tree max depth	25	15	25	25	15	15	50	15	15	15
Tree minimum child weight	10	1	5	5	5	1	10	1	10	10
Gamma	0	0.01	0.001	0.01	0	0.01	0	0	0.001	0.01
Subsample ratio	1	0.7	0.7	0.7	0.6	0.6	0.7	0.8	1	1
Columns sampling by tree	0.7	1	1	0.8	1	1	1	1	0.7	0.7
Alpha	0	1	0	0	0.1	0	0.01	1	1	0.1
Learning rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1	0.1	1

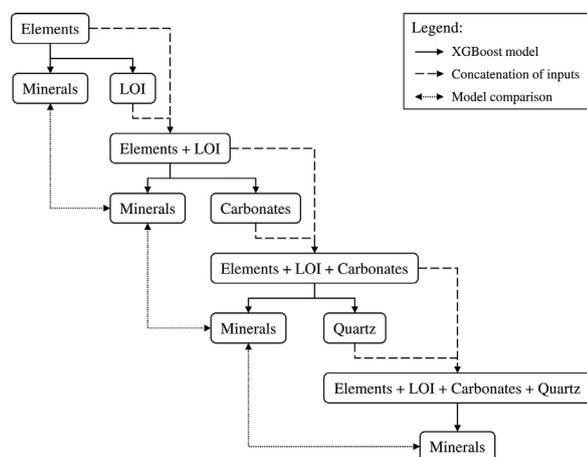


Fig. 4. SML workflow proposed in the present research. Solid arrows indicate the application of the XGBoost model to estimate minerals and LOI, while dashed arrows indicate the concatenation of the modeled output with the previous inputs. Dotted arrows indicate a comparison between estimates from different steps.

Clay model: due to the inherent challenges of estimating clay minerals, quartz was added to the input data, and a new model was created for clay. It is expected that a model created using the chemical elements + LOI + carbonates + quartz will be able to aggregate not only information on the composition but also insights into the clay's depositional environment, achieving better results. As in previous training, models for other minerals were created to be compared with previous results.

Fig. 4 presents a summary of the stepped learning proposed in this research.

Evaluation and test

R^2 and MSE values of the validation and cross-validation with and without SML were compared to determine the best models. Uncertainty of the models, measured from the standard deviation of the error, was also used to evaluate the SML. In the test phase, the trained models were applied to three wells drilled in the pre-salt, using their geochemical logs. The mineralogy acquired was compared to XRD analyses and used to make considerations about the wells' rocks.

Results

Stepped machine learning models

The results of the SML are presented in [Table 5](#). As expected, a clear improvement in R^2 and a decrease in MSE is observed as new information is added to the models' input data.

First, the LOI model's quality is observed, with R^2 of 0.962 and MSE of 0.00028 in the validation data and R^2 of 0.956 and MSE of 0.00041 in cross-validation. These results give greater confidence in the use of LOI in subsequent models. The inclusion of LOI improves the carbonate model, both in the validation and cross-validation results. The inclusion of carbonates in the input data generates a substantial improvement in calcite, dolomite, and quartz predictions. For quartz, R^2 increases from values below 0.70 to above 0.85, and MSE decreases to less than half for both validation and cross-validation. Finally, the inclusion of quartz in the inputs generated a significant improvement in the clay model, with R^2 greater than 0.80 and MSE decreasing to almost a third for validation and cross-validation. The inclusion of quartz did not significantly improve the calcite and dolomite models.

The K-feldspar, plagioclase, and pyroxene models showed R^2 and MSE relatively constant, regardless of SML. The pyroxene model showed the worst results overall, with R^2 below 0.80. However, the model for plagioclase + pyroxene using chemical elements, LOI, and carbonates as input presented better results than pyroxene alone.

[Table 6](#) shows the evolution of the standard deviation of errors in the models generated by stepped learning. The standard deviation of the error can be understood as a measure of the models' uncertainty since it indicates how far the predicted value is from the real value. In general, uncertainty decreases as more variables are added to the input data. This phenomenon is expected since it is a reflection of the increase in R^2 of the models. [Fig. 5](#) shows real versus modeled data for LOI, carbonates, K-feldspar, and plagioclase. [Figs. 6 and 7](#) show real versus modeled data for the other minerals before and after the SML application to illustrate the model's improvement. A significant decrease in the points' dispersion is observed.

Variable importance

[Fig. 8](#) shows the importance of the variables during the training of the mineralogical model. The importance presented is related to the latest models created for each mineral and LOI; therefore, some amounts will be null (e.g., LOI in the K-feldspar model, since it was not used).

Table 5
R² results of the SML models.

	Validation set				Cross-validation			
	Elements	Elements + LOI	Elements + LOI + Carb	Elements + LOI + Carb + Qtz	Elements	Elements + LOI	Elements + LOI + Carb	Elements + LOI + Carb + Qtz
	R²							
LOI	0.962	-	-	-	0.956	-	-	-
Carb	0.815	0.845	-	-	0.821	0.852	-	-
Cal	0.758	0.780	0.862	0.874	0.755	0.768	0.840	0.850
Dol	0.781	0.816	0.823	0.850	0.730	0.769	0.774	0.790
Qtz	0.673	0.663	0.869	-	0.672	0.683	0.850	-
Kfd	0.815	0.822	0.811	0.815	0.708	0.701	0.708	0.698
Clay	0.597	0.662	0.698	0.853	0.637	0.687	0.717	0.814
Plg	0.960	0.964	0.962	0.962	0.636	0.637	0.653	0.644
Prx	0.632	0.638	0.674	0.674	0.788	0.784	0.786	0.783
Plg + Prx	0.860	0.880	0.914	0.933	0.813	0.828	0.825	0.795
	MSE							
LOI	2.8E-04	-	-	-	4.1E-04	-	-	-
Carb	8.8E-03	7.4E-03	-	-	7.7E-03	6.5E-03	-	-
Cal	1.6E-02	1.4E-02	8.9E-03	8.1E-03	1.5E-02	1.4E-02	1.0E-02	9.3E-03
Dol	1.1E-02	8.8E-03	8.5E-03	7.2E-03	1.1E-02	9.6E-03	9.4E-03	8.6E-03
Qtz	6.9E-03	7.1E-03	2.8E-03	-	6.1E-03	5.8E-03	2.7E-03	-
Kfd	4.2E-04	4.0E-04	4.3E-04	4.2E-04	6.5E-04	6.7E-04	6.4E-04	6.4E-04
Clay	2.0E-03	1.7E-03	1.5E-03	7.4E-04	2.1E-03	1.8E-03	1.7E-03	1.1E-03
Plg	2.6E-04	2.4E-04	2.5E-04	2.5E-04	1.5E-03	1.4E-03	1.3E-03	1.4E-03
Prx	4.9E-04	4.8E-04	4.4E-04	4.4E-04	2.1E-04	2.2E-04	2.1E-04	2.1E-04
Plg + Prx	1.8E-03	1.5E-03	1.1E-03	8.6E-04	1.2E-03	1.3E-03	1.4E-03	1.4E-03

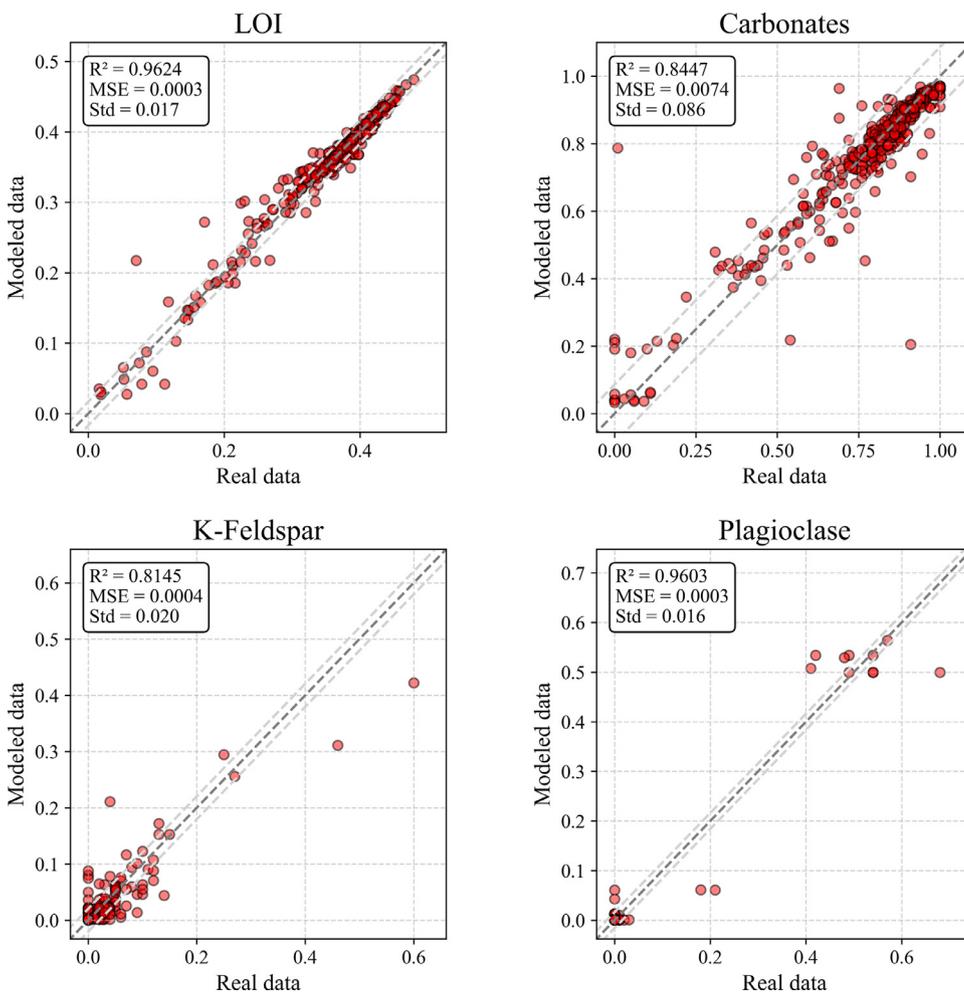


Fig. 5. Real data versus modeled data graphs for the test sets of LOI, carbonates, K-feldspar, and plagioclase.

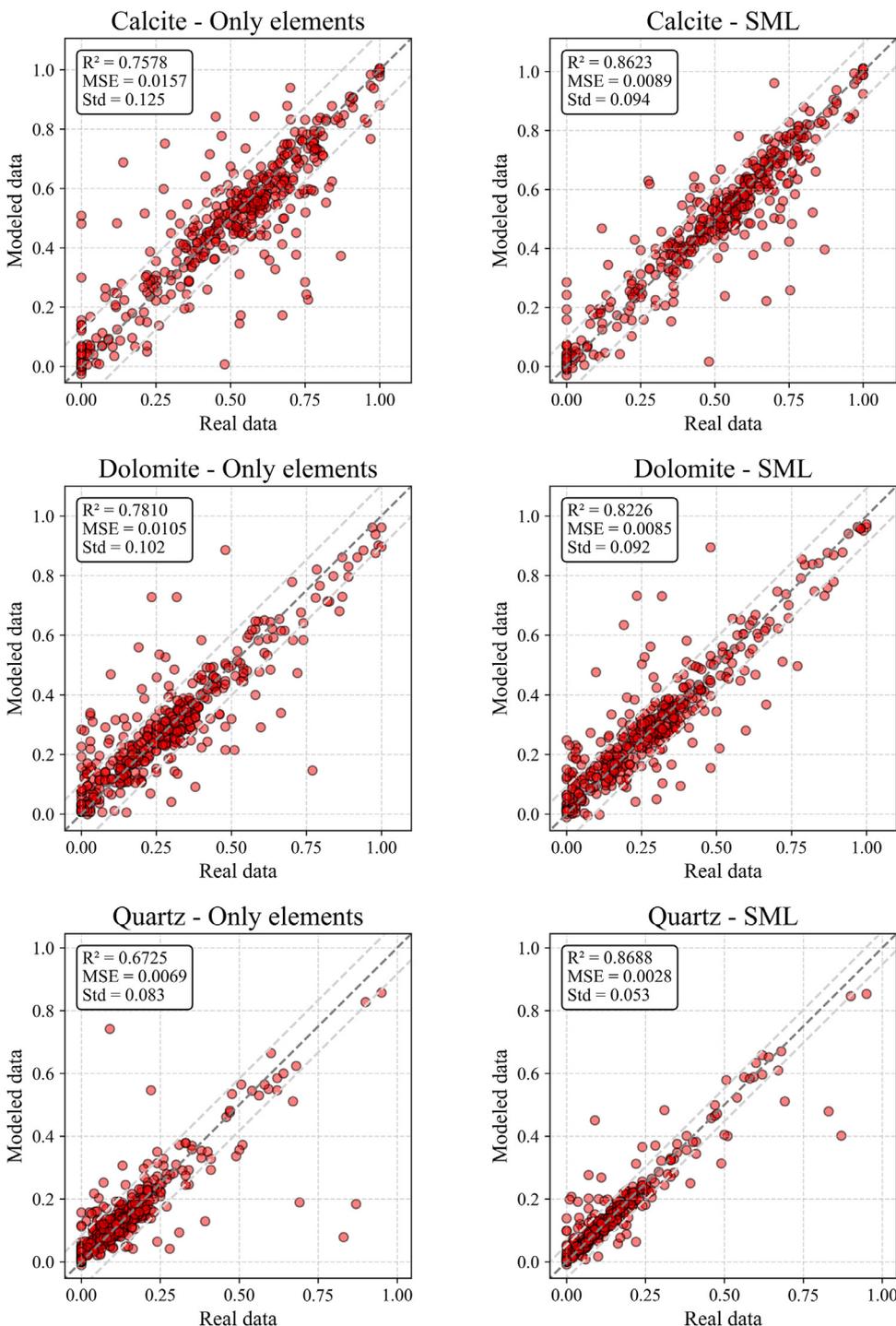


Fig. 6. Real data versus modeled data graphs for the test sets of calcite, dolomite, and quartz before and after stepped learning.

Ca and Si were the most critical elements for the LOI model. The importance of Ca is due to the high correlation of LOI with carbonates, whose main element is calcium. Si's importance is probably due to the replacement of carbonate minerals with siliciclastic and igneous minerals rich in silicon, a typical pre-salt pattern. As expected, the carbonate model (calcite + dolomite) is highly impacted by LOI. The calcite model is impacted by Ca, Mg, and carbonates, and the dolomite model is impacted by Mg, LOI, and Ca, something expected for pre-salt carbonates. Na's importance in the dolomite model can be explained by Na and Mg's presence in plagioclase and pyroxene. The increase in Mg together with the increase in Na may indicate the presence of plagioclase and pyroxene, whereas the increase in Mg alone

means the presence of dolomite, making Na important in the dolomite model.

The quartz model is mainly impacted by Si and carbonates, reflecting carbonates' replacement by siliciclastic/igneous minerals. The elements K and Al are the most important in the K-feldspar model because they are present in its chemical composition. Although not present in K-feldspar, Ca and Fe are possibly useful to differentiate plagioclase and pyroxene. The clay model is influenced mainly by Mg, Na, LOI, carbonates, quartz, and secondarily by Si and Ca, reflecting both the varied chemical composition of the pre-salt clays and their depositional environment. This result demonstrates how complex it would be to create a clay prediction model without using SML.

Table 6
Standard deviations of the errors obtained in the validation sets of SML.

	Global Error Std (g/g)			
	Elements	Elements	Elements	Elements
	Elements + LOI	+ LOI + Carb	+ LOI + Carb + Qtz	
LOI	0.027	-	-	-
Carb	0.095	0.087	-	-
Cal	0.121	0.120	0.096	0.094
Dol	0.100	0.090	0.088	0.087
Qtz	0.085	0.085	0.051	-
Kfd	0.021	0.021	0.021	0.021
Clay	0.045	0.041	0.039	0.028
Plg	0.016	0.018	0.017	0.017
Prx	0.022	0.022	0.021	0.021
Plg + Prx	0.046	0.046	0.042	0.041

The plagioclase and pyroxene + plagioclase models are impacted by Al, an element present in plagioclase composition. Ti, Al, Fe, and Na are essential in the pyroxene model. Al and Na are probably related to the high correlation between pyroxene and plagioclase. Ti and Fe may be present in the pyroxene family’s minerals, demonstrating the complexity of creating a traditional mineral model for pre-salt rocks.

Test

The models generated by SML were applied to the geochemical data of three wells drilled in the pre-salt, and the results are shown in Figs. 9, 10, and 11. These figures also present the estimated mineralogy using

only the chemical elements, without the application of SML. The mineralogy logs were compared with the XRD analyzes of rotary sidewall cores and core plugs collected in these wells. As these samples do not have XRF analyzes, they were not used to train SML models.

Carbonates of the Barra Velha Formation are observed in Well A, formed predominantly by calcite, dolomite, and quartz. The concentrations of these minerals estimated by the SML model agree with that observed in the XRD analyzes of the rock samples. From X020 m, a decrease in the calcite concentration is observed, detected by the model and XRD analyzes. In the final depths, a layer rich in dolomite, quartz, clay, and K-feldspar is observed, related to the Piçarras Formation’s siliciclastic minerals. The SML model can correctly estimate the concentration of these minerals. Regarding the estimated mineralogy using only the chemical elements, pyroxene has anomalous concentrations over the entire range, a problem corrected by the SML model.

In Well B, the Barra Velha and Itapema Formations are observed. The carbonates of the Barra Velha Formation have higher concentrations of dolomite and quartz, while the carbonates of the Itapema Formation are formed mainly by calcite. These characteristics are captured by the SML model and confirmed by the XRD analysis of the rock samples. Between the two carbonates (X660 to X670 m), a layer of siliciclastic rock is marked by the increased concentration of clay and K-feldspar in XRD analyzes. The SML model was able to identify these concentrations correctly. Comparing the mineralogy estimated by SML and only by the chemical elements, the same anomaly in the pyroxene concentration seen in Well A is also observed. Also, the concentrations of clay and dolomite show noisier values when estimated only by the chemical elements, possibly reflecting the worse stability of these models compared to that obtained by SML.

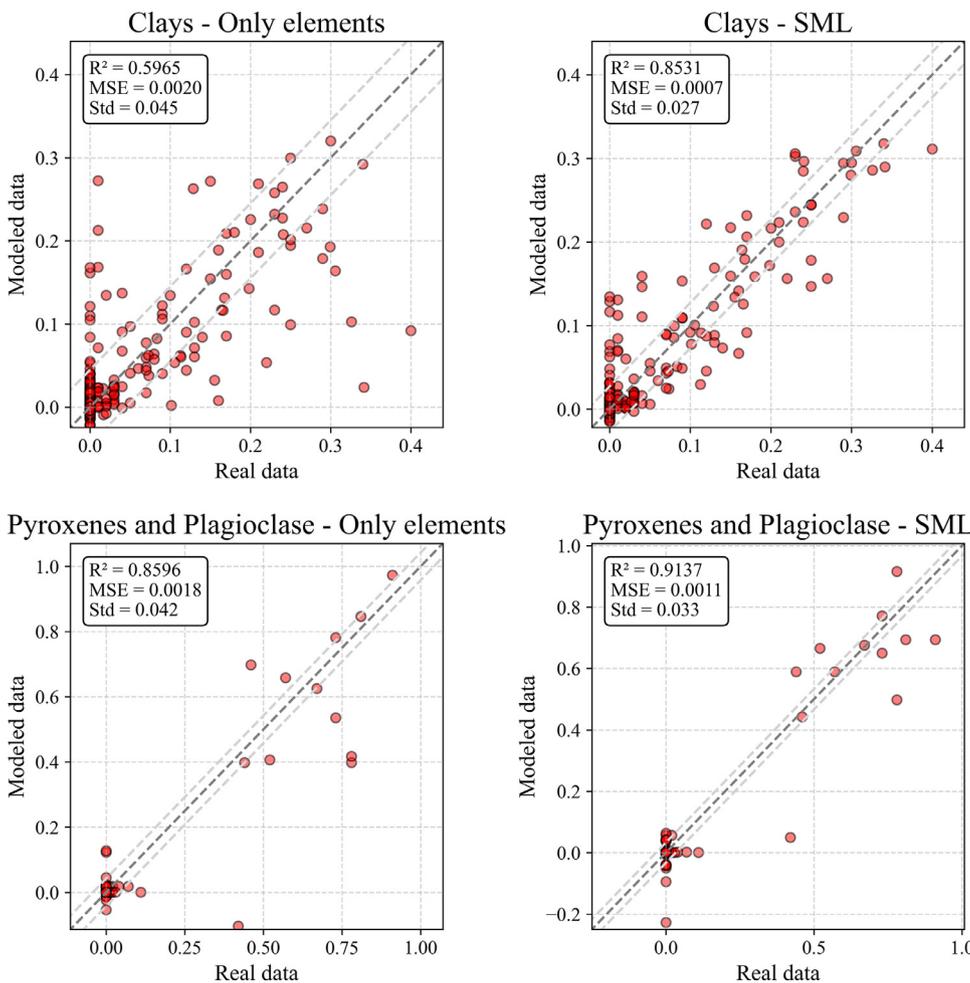


Fig. 7. Real data versus modeled data graphs for the test sets of clays and pyroxenes + plagioclase before and after stepped learning.

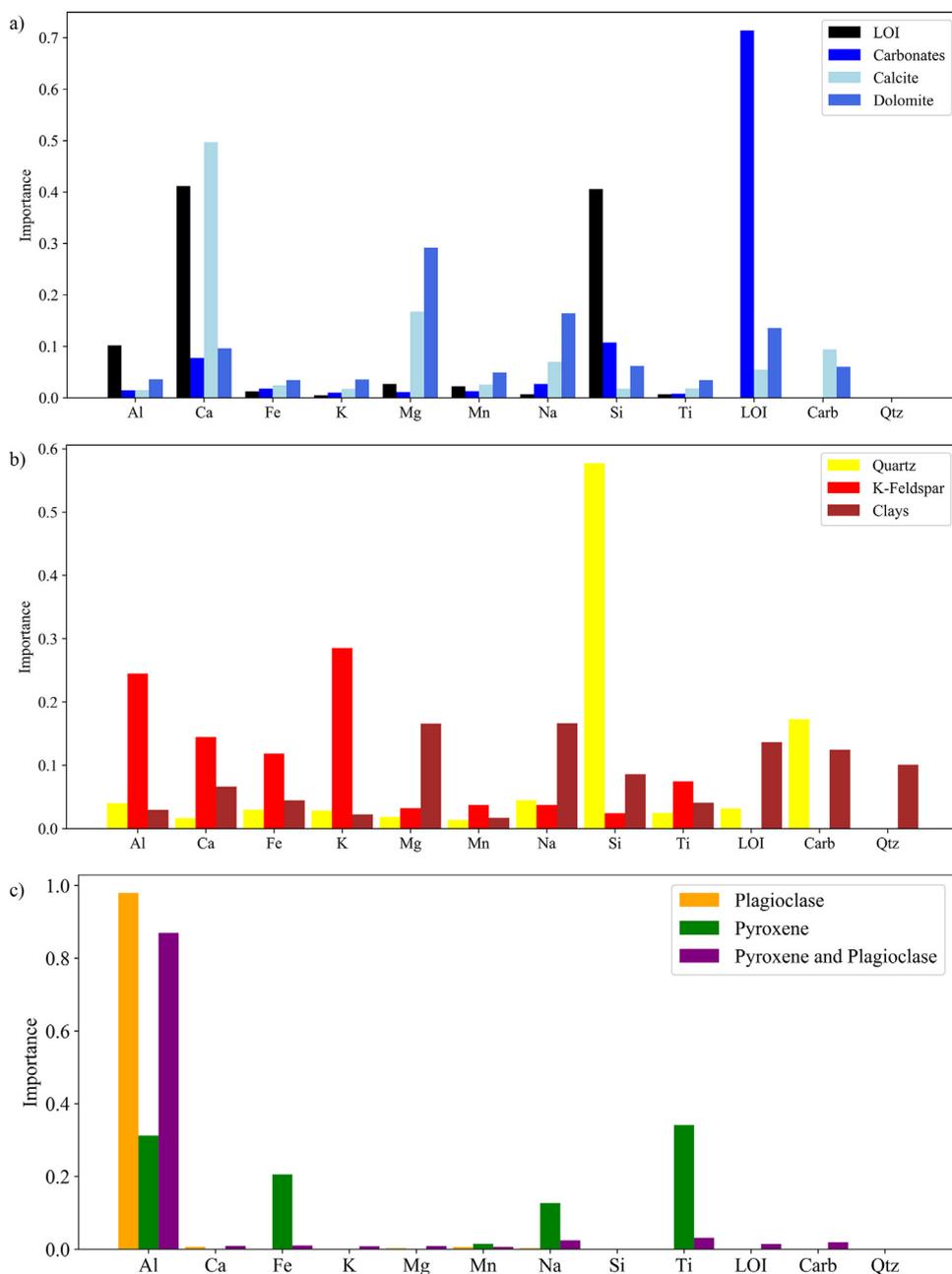


Fig. 8. Variable importance of the chemical elements during the training of the mineralogical model. a) The importance of the variables of the LOI, carbonate, calcite, and dolomite models. b) The importance of the variables of the quartz, K-feldspar, and clay models. c) The importance of the variables of the plagioclase, pyroxene, and pyroxene + plagioclase models.

A thick layer of igneous rock is observed in the first half of Well C. This rock's presence is marked by the increase in plagioclase and pyroxene observed in XRD analyzes of some rock samples. The SML model correctly identified these concentrations. In the rest of the well, the SML model correctly estimated the concentrations of calcite, dolomite, and quartz present in carbonates. Again, anomalous pyroxene concentrations are observed in the mineralogy estimated only by the chemical elements.

It is important to note that the logs acquired by wireline tools have a vertical resolution much lower than that of a core sample. Therefore, the comparison between these two pieces of information should not focus on exact values but general trends. Even so, the concentrations obtained by the SML model were very close to those observed in the core samples.

Discussion

The stepped training strategy generated better results for mineralogical estimates than the individual training of mineral models, marked

by the improvement of R^2 in the comparison between real and modeled data and the decrease of the standard deviation of the models' error. Since the standard deviation of the error indicates how wrong a model can be, it can be understood as the model's uncertainty. As the models are not perfect and always have some error, it is essential to emphasize that the stepped training will propagate a model's uncertainties to the others. Therefore, the ideal is that step training is used to promote a clear improvement of the models.

After analyzing the evolution of R^2 and the uncertainty, the best SML flow was defined to create the mineralogical model for the pre-salt rocks. It is shown in Fig. 12, summarized below.

- Estimation of LOI, K-feldspar, and plagioclase using only the chemical elements.
- Inclusion of LOI to chemical elements for the estimation of carbonates (calcite + dolomite).
- Inclusion of carbonates in the chemical elements and LOI for the estimation of calcite, dolomite, quartz, and pyroxene + plagioclase.

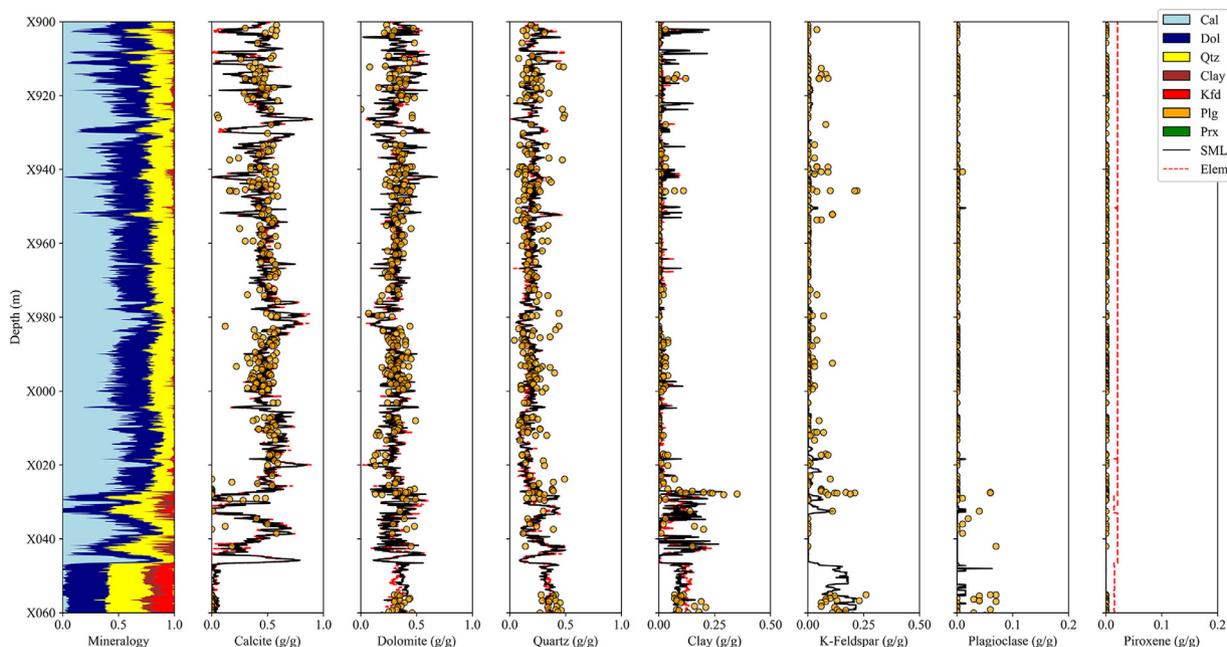


Fig. 9. Mineralogy obtained by applying the SML model to the geochemical logs of Well A and comparison with the mineralogy obtained using only the chemical elements (Elem, dashed red) and XRD analyzes of rock samples. The SML model honors calcite, dolomite, and quartz concentrations in the well’s initial depths. The model was also able to correctly estimate the increase in the concentrations of dolomite, quartz, clay, and K-feldspar observed at the end of the well. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

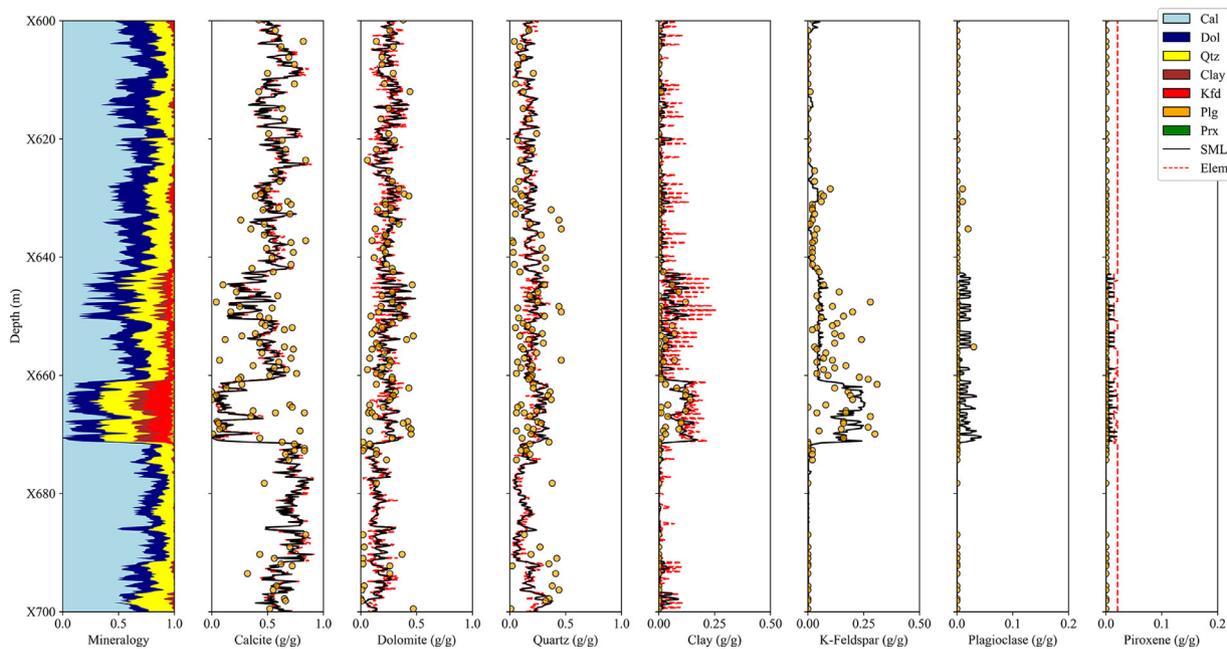


Fig. 10. Mineralogy obtained by applying the SML model to the geochemical logs of Well B and comparison with the mineralogy obtained using only the chemical elements (Elem, dashed red) XRD analyzes of rock samples. The SML model was able to detect the increase in clay and K-feldspar observed in the middle of the well. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Subtraction of plagioclase from the pyroxene + plagioclase for the estimation of pyroxene.
- Inclusion of quartz to chemical elements, LOI, and carbonates for the estimation of clay.

With the proposed SML flow, the best results were extracted from each mineral’s models with minimal propagation of uncertainties. The SML models were tested in geochemical logs of wells drilled in the pre-salt rocks, demonstrating the strategy’s robustness.

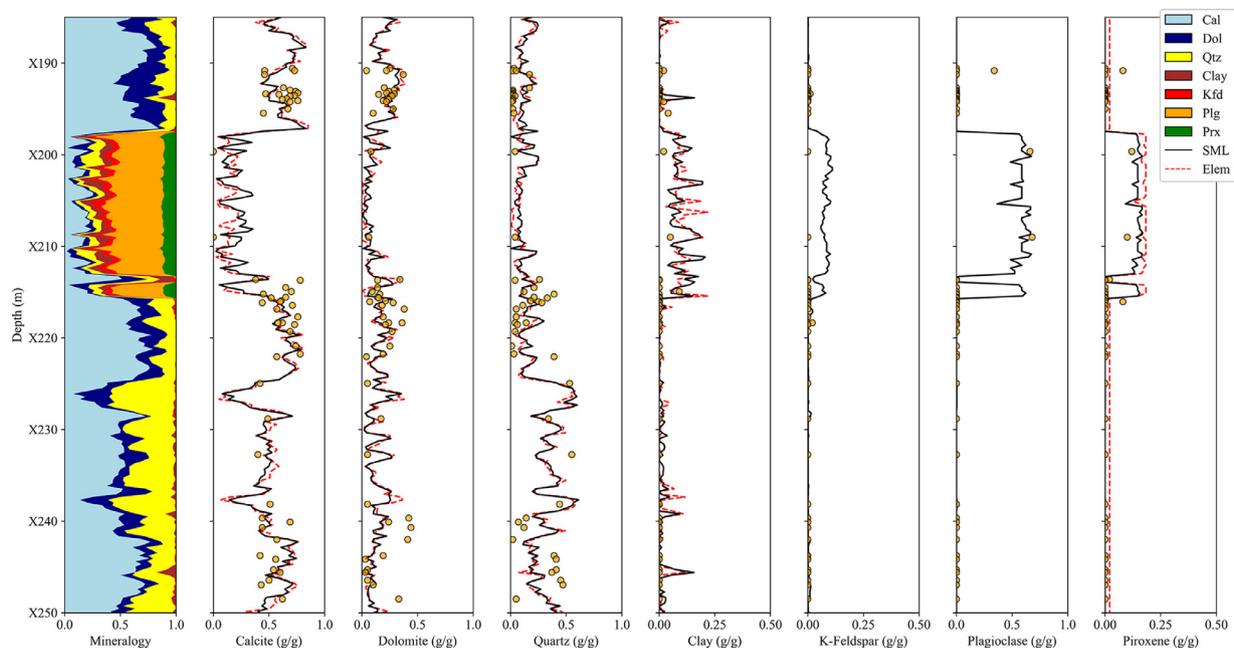


Fig. 11. Mineralogy obtained by applying the SML model to the geochemical logs of Well C and comparison with the mineralogy obtained using only the chemical elements (Elem, dashed red) XRD analyzes of rock samples. The SML model correctly detected the igneous rock layer observed in the first half of the well, with high plagioclase and pyroxene concentrations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

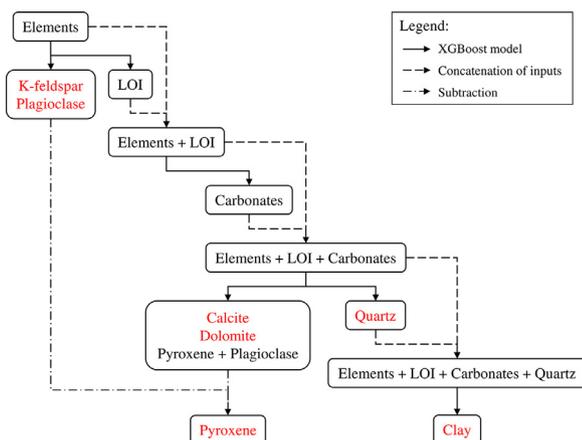


Fig. 12. The SML proposed for the mineral model for pre-salt rocks. Solid arrows indicate the application of the XGBoost model to estimate minerals and LOI, while dashed arrows indicate the concatenation of the modeled output with the previous inputs. Dash-dotted arrows indicate the subtraction of the concentration of pyroxene + plagioclase by plagioclase to obtain pyroxene.

Conclusion

Four machine learning algorithms were tested to create a mineralogical model for pre-salt rocks: MLP, GAN, Random Forest, and XGBoost. XGBoost presented the best initial results, being the chosen algorithm for the generation of mineral models.

The stepped training proposed in the present research generated significant improvements in the mineral models' quality compared to the model using only the chemical elements. The SML model increased the R^2 of the carbonate model from 0.815 to 0.845, of calcite from 0.758 to 0.862, of the dolomite model from 0.781 to 0.823, of quartz from 0.673 to 0.869, and of the clay model from 0.597 to 0.853. A decrease in MSE was also observed.

The application of stepped training propagates the errors and uncertainties of a model. Thus, the SML strategy was not applied to K-

feldspar, plagioclase, and pyroxene since the stepped training did not significantly improve the quality of these minerals' models, ensuring that errors were not propagated unnecessarily.

The SML model estimated the mineralogy of three wells drilled in the pre-salt using the geochemical logs. Samples from these wells were not used in training the models. The results obtained honored the XRD analyzes of the samples from these wells, identifying calcitic carbonates with varying concentrations of dolomite and quartz, siliciclastic rocks rich in clay and K-feldspar, and igneous rocks composed of plagioclase and pyroxene. These analyses attested to the model's ability to perform well in the different scenarios found in the pre-salt.

The present research demonstrated that the integration between the machine learning tools and the geological knowledge in stepped learning was crucial for creating a mineralogical model for the Brazilian pre-salt rocks. The real gain of the machine learning application comes not only from its use alone but from what it can add to the previously existing geological knowledge. It is expected that the application of SML can be expanded to solve new challenges found in the pre-salt and other hydrocarbon reservoirs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to the Brazilian National Agency for Petroleum, Natural Gas and Biofuels (ANP), and Petrobras for supporting this research, and LCT Laboratory for performing part of the analysis. This research was developed with the RDI Group Integrated Technology of Rock and Fluid Analysis (InTRA) and conducted under the Postgraduate Program from the Naval and Oceanic Engineering Department of the Escola Politécnica, Universidade de São Paulo.

References

- [1] Aggarwal K, Kirchmeyer M, Yadav P, Keerthi SS, Gallinari P. Benchmarking regression methods: a comparison with CGAN; 2019. Retrieved from <http://arxiv.org/abs/1905.12868>.
- [2] Anderson RN, Dove RE, Boglia C, Silver LT, James EW, Chappell BW. Elemental and mineralogical analyses using geochemical logs from the Cajon Pass scientific drillhole, California, and their preliminary comparison with core analyses. *Geophys Res Lett* 1988;15(9):969–72.
- [3] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32. doi:10.1023/A:1010933404324.
- [4] Breiman L, Cutler A, Liaw A, Wiener M. Breiman and Cutler's Random Forests for classification and regression; 2018. Retrieved from <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>.
- [5] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: 22nd KDD conference on knowledge discovery and data mining; 2016. p. 785–94. doi:10.1145/2939672.2939785.
- [6] Cullity B, Stock S. Elements of X-ray diffraction. 3rd ed. New York: Prentice-Hall; 2001.
- [7] Dean WEJ. Determination of carbonate and organic matter in calcareous sediments and sedimentary rocks by loss on ignition: comparison with other methods. *J Sediment Petrol* 1974;44:242–8.
- [8] Ellis DV, Singer JM. Well logging for earth scientists. Second. Springer; 2007.
- [9] Flaum C, Pirie G. Determination of lithology from induced gamma ray spectroscopy. *SPWLA 23rd Annual Logging Symposium*, 16. Mexico City. Mexico: Society of Petrophysicists and Well-Log Analysts 1981.
- [10] Freedman R, Herron S, Anand V, Herron M, May D, Rose D. New method for determining mineralogy and matrix properties from elemental chemistry measured by gamma ray spectroscopy logging tools. *SPE Reserv Evaluat Eng* 2015;18(04):599–608. doi:10.2118/170722-pa.
- [11] Freund Y, Schapire RE. Large margin classification using the perceptron algorithm. *Mach Learn* 1999;37(3):277–96. doi:10.1023/A:1007662407062.
- [12] Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;39.
- [13] Gonzalez J, Lewis R, Hemingway J, Grau J, Rylander E, Schmitt R. Determination of formation organic carbon content using a new neutron-induced gamma ray spectroscopy service that directly measures carbon. *SPWLA 54th annual logging symposium*, 15; 2013. doi:10.1190/urtec2013-112.
- [14] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Bengio Y. Generative adversarial nets. *Adv Neural Inf Process Syst* 2014;3:2672–80.
- [15] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. In *springer series in statistics*. Second. Springer; 2009.
- [16] Heiri O, Lotter AF, Lemcke G. Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results. *J Paleolimnol* 2001;25:101–10. doi:10.1023/A:1008119611481E.
- [17] Herron MM. Mineralogy from geochemical well logging. *Clays Clay Min* 1986;34(2):204–13. doi:10.1346/ccmn.1986.0340211.
- [18] Herron S, Herron M, Pirie I, Saldungaray P, Craddock P, Charsky A, Li T. Application and quality control of core data for the development and validation of elemental spectroscopy log interpretation. *SPWLA 55th annual logging symposium*, 23. Abu Dhabi: UAE: Society of Petrophysicists and Well-Log Analysts; 2014.
- [19] Ho TK. Random decision forests. In: Proceedings of the international conference on document analysis and recognition, ICAR; 1995. p. 278–82. doi:10.1109/IC-DAR.1995.598994.
- [20] Jenkins R, Snyder RL. Introduction to X-ray powder diffractometry; 1996. doi:10.1002/9781118520994.
- [21] Jin X, Zhu D, Hill AD, McDuff D. Effects of heterogeneity in mineralogy distribution on acid fracturing efficiency. In: Society of petroleum engineers - hydraulic fracturing technology conference and exhibition; 2019. p. 147–60. doi:10.2118/194377-pa.
- [22] Kuhn M, Johnson K. Applied predictive modeling. Springer; 2013. doi:10.1007/978-1-4614-6849-3.
- [23] Moreira JLP, Valdetaro C, Gil JA, Machado MAP. Bacia de Santos. *Boletim de Geociencias Da Petrobras* 2007;15(2):531–49.
- [24] North RJ. Through-casing reservoir evaluation using gamma ray spectroscopy. In: SPE california regional meeting. Ventura, California, USA: Society of Petroleum Engineers; 1987. p. 329–42.
- [25] Oliveira LAB, Carneiro Cde C. Synthetic geochemical well logs generation using ensemble machine learning technics for the Brazilian pre-salt reservoirs. *J Petrol Sci Eng* 2021;196:24 January 2021. doi:10.1016/j.petrol.2020.108080.
- [26] Rosenblatt F. The Perceptron - A perceiving and recognizing automaton. In *Report*, 85. Buffalo, NY: Cornell Aeronautical Laboratory; 1957.
- [27] Simon AH. Sputter processing. In: Seshan K, Schepis D, editors. Handbook of Thin Film Deposition; 2018. p. 195–230. <https://doi.org/https://doi.org/doi:10.1016/C2016-0-03243-6>.
- [28] Ulloa JM, Chaparro D, Lara S, Arango S, Mendez F, Alarcon N, Gade S. An innovative cased-hole, oil-saturation method of utilizing excess carbon analysis of pulsed neutron measurements in a siliciclastic cenozoic formation, Los Llanos Basin, Colombia.. *SPWLA 57th annual logging symposium*, 12. Reykjavik, Iceland: society of petrophysicists and well-log analysts; 2016.
- [29] Wendemuth A. Learning the unlearnable. *J Phys A: Math Gen* 1995;28:5423–36. doi:10.1088/0305-4470/28/18/030.
- [30] Westaway P, Hertzog R, Plasek RE. Neutron-induced gamma ray spectroscopy for reservoir analysis. *Soc Petrol Eng J* 1983;23(03):553–64. doi:10.2118/9461-pa.
- [31] XGBoost documentation; 2015. Retrieved June 52019 from <https://xgboost.readthedocs.io/en/latest/parameter.html>.
- [32] Zhao J, Chen H, Yin L, Li N. Mineral inversion for element capture spectroscopy logging based on optimization theory. *J Geophys Eng* 2017;14(6):1430–6. doi:10.1088/1742-2140/aa7bfa.