

# 11º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS



## TÍTULO DO TRABALHO:

SIMILARIDADE ENTRE PERFIS DE POÇOS DE PETRÓLEO: UMA ABORDAGEM A PARTIR  
DE CIÊNCIA DE DADOS E MAPAS AUTO-ORGANIZÁVEIS

## AUTORES:

Cleyton de Carvalho Carneiro<sup>1</sup>, Rodrigo César Teixeira de Gouvêa<sup>1</sup>, Rafael dos Santos Gioria<sup>1</sup>, Afonso Paiva Neto<sup>2</sup>

## INSTITUIÇÃO:

<sup>1</sup> Interações Tecnológicas em Análises de Rochas e Fluidos (InTRA), Departamento de Engenharia de Minas e de Petróleo, Escola Politécnica, Universidade de São Paulo – Campus Santos (PMI, EP, USP);

<sup>2</sup> Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo – Campus São Carlos (SME, ICMC, USP).

**SIMILARIDADE ENTRE PERFIS DE POÇOS DE PETRÓLEO: UMA ABORDAGEM A PARTIR DE CIÊNCIA DE DADOS E MAPAS AUTO-ORGANIZÁVEIS**

**Abstract**

The selection of similar wells is a recurring demand for obtaining strategic information in research related to oil reservoirs. However, the formulation of a measure of similarity between wells requires the development of a particular systematic. Well logs can be analyzed as a time series, replacing the reference variable time with depth. The growing development of data science and machine learning algorithms has enabled the use of self-organizing applications on several research fronts. However, the self-organization of time series still faces limitations, mainly inherent to the topology of the organization systems. In 2021, the Brazilian National Agency of Petroleum, Natural Gas, and Biofuels (ANP) made all wells in the Brazilian continental basins available free of charge for research. The research presented here starts from the development of self-organized series from the basic suite of oil well profiles, seeking to reduce the dimensionality and compare the similarity between these series. A new methodological approach was developed based on the Self-Organizing Maps (SOM) algorithm, comprising the creation of a position matrix, called P-Matrix, which was analyzed together with the D-Matrix, obtained from the redistribution of the U-Matrix distance values. In addition, a new matrix called IDP-Matrix was developed based on an index that integrates distance and position. The newly generated matrices allowed the reconstruction of the time series after the self-organization process, keeping topological relationships of the n-dimensional space. This process was named here as Self-Organizing Series (SOrS). Therefore, the series based on position (P-SOrS), distance (D-SOrS), and position and distance index (IDP-SOrS) were defined. The signatures were compared and ordered by similarity according to the nonlinear alignment cost calculated by the Dynamic Time Warping (DTW) algorithm. SOrS were generated for 30 wells from 5 Brazilian continental sedimentary basins. The series obtained preserved the lithostratigraphic characteristics observed in the composite profiles. The similarity ranking graphs developed for each sedimentary basin showed evident stratigraphic similarities, such as in the Parnaíba and Solimões basins. The SOrS, therefore, proved to be a viable and useful alternative to the well similarity system, maintaining intrinsic characteristics of the original profiles. The methodology proves to be efficient and scalable for application in other research fronts and mining the oil wells' databases on the ocean platform.

**Introdução**

Algoritmos de inteligência artificial têm sido empregados para suprir diversos desafios na avaliação de reservatórios, sobretudo na fase de produção dos campos. Dada a crescente demanda por extensas bases de dados, esses algoritmos têm se demonstrado como ferramental eficiente para treinar modelos utilizando as informações adquiridas em poços perfurados na fase exploratória e no início da fase de produção.

Espera-se que o aprendizado de máquina aplicado às extensas bases de dados existentes de centenas de campos produtores permita uma otimização sem precedentes na aquisição de informações. Entretanto, para que os melhores resultados sejam atingidos, é desejável que esses algoritmos sejam treinados utilizando dados representativos dos novos poços que são perfurados. Modelos de aprendizado de máquinas treinados indiscriminadamente podem gerar resultados pouco representativos dos poços perfurados, principalmente durante a fase exploratória.

## **11º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS**

Sendo assim, é necessário que seja criada uma etapa anterior à aplicação de algoritmos de aprendizado de máquina aos dados adquiridos em novos poços. Essa etapa consiste na comparação dos dados adquiridos inicialmente em um novo poço a uma base de dados de poços perfurados em diversos reservatórios, obtendo-se assim uma medida de similaridade entre poços. De posse da medida de similaridade seria possível selecionar da base de dados os poços mais representativos e utilizá-los para treinar modelos que seriam aplicados ao novo poço. A depender do grau de similaridade observado, a aquisição de dados do novo poço poderia ser substituída pelos dados sintéticos, reduzindo custos e tornando as fases exploratórias e de produção mais eficientes. Essa medida de similaridade poderia ser construída, em um futuro, levando em consideração informações como os dados sísmicos e alguns perfis de poços, e o grau de similaridade poderia ser usado para suprimir a aquisição de perfis de poços de mais alto custo, amostras de rocha e fluido, e testes de formação, permitindo a antecipação de informações relativas à produção esperada para o poço utilizando apenas um pequeno conjunto de informações. A abrangência do cálculo da similaridade entre poços permitirá uma comparação inicial de diferentes poços.

Embora os estudos da similaridade possuam amplas aplicações em diversas áreas de pesquisa, a literatura científica e tecnológica ainda possui uma lacuna em relação aos estudos de similaridade entre os poços de petróleo. Os trabalhos de (DU et al., 2020) abordam a medida de similaridade baseada em características morfológicas na análise de perfis de poços em arenitos. No entanto, o trabalho em questão se utiliza de análises em uma única variável de raios gamma, o que torna a comparação bem mais simples do que na abordagem multivariada. A similaridade entre perfis de poços também foi descrita por ALI et al., (2021) em uma abordagem para a predição de dados faltantes em perfis sônicos com o uso da similaridade de perfis. A similaridade, neste caso, foi calculada com o uso de redes neurais profundas, do inglês, Deep Neural Networks (DNN) em conjunto com métricas das similaridades de Jaccard e Overlap.

A peculiaridade das informações obtidas pelos perfis de poços e a alta quantidade de informação disponível, inclusive publicamente, se somam à vasta aplicação destes perfis na indústria do petróleo. A ciência de dados e a inteligência artificial, a partir de suas abordagens, possibilitam a utilização da informação contida nos perfis de poços buscando, a partir destes dados, compreender relações de similaridade entre poços. A similaridade de poços, por sua vez, poderá subsidiar a escolha de parâmetros de treinamento para aprendizado de máquina na construção de inúmeros modelos para a indústria do petróleo. Tais modelos orientam a tomada de decisão e minimizam custos operacionais.

Este trabalho se propõe a analisar a similaridade entre poços de petróleo a partir da redução de dimensionalidade gerada por uma suíte básica de perfis de poços, composta por cinco variáveis.

## **Metodologia**

### **Aquisição da Base de dados de Poços de Exploração e Produção de Petróleo**

A Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) disponibilizou no ano de 2021 uma ampla base de dados com informações de exploração e produção em bacias sedimentares continentais brasileiras. Para esta pesquisa foram selecionados 30 poços de 10 campos de produção relativos a 5 bacias sedimentares. Os poços selecionados estão listados na Tabela 1. Os poços selecionados possuem posição espacial e profundidades distintas. Em relação às variáveis de perfilagem de poços medidas, foram selecionadas cinco perfis, a saber: (i) raios gama; (ii) densidade; (iii) nêutrons; (iv) fator fotoelétrico; e (v) sônico. Cada profundidade de aquisição em cada um dos poços foi tratada

## 11º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS

como uma instância na base de dados. A base de dados final reuniu 376.993 instâncias relacionadas aos registros das 5 variáveis de perfilagem em diferentes profundidades dos 30 poços, empilhadas sequencialmente.

Tabela 1 – Poços selecionados para compor a base de dados e seus respectivos campos de produção e bacias sedimentares, bem como a profundidade de aquisição dos perfis

Poço	Campo de Produção	Bacia Sedimentar	Profundidade (m)
3-BRSA-1148-SE	Siririzinho	Sergipe	609
3-BRSA-1098-SE	Siririzinho	Sergipe	760
3-BRSA-1093-SE	Siririzinho	Sergipe	886,7
3-SZ-464-SE	Siririzinho	Sergipe	572,2
3-BRSA-912-SE	Carmópolis	Sergipe	246,3
3-BRSA-1279-SE	Carmópolis	Sergipe	723,5
3-BRSA-1262-SE	Carmópolis	Sergipe	839,5
3-BRSA-1307D-SE	Carmópolis	Sergipe	669,2
6-BRSA-1202D-BA	Dom João	Recôncavo	1427,5
1-BRSA-470D-BA	Guanambi	Recôncavo	921,1
4-BRSA-422-BA	CEXIS	Recôncavo	1810
3-BRSA-349-RN	Canto do Amaro	Potiguar	1794,4
3-BRSA-1295-RN	Canto do Amaro	Potiguar	1065,3
3-BRSA-1193-RN	Canto do Amaro	Potiguar	948,3
3-BRSA-1224-RN	Estreito	Potiguar	437,8
3-BRSA-1210-RN	Estreito	Potiguar	1492,8
3-BRSA-1325-RN	Estreito	Potiguar	492,9
3-BRSA-1320-RN	Alto do Rodrigues	Potiguar	483,3
3-BRSA-1341-RN	Alto do Rodrigues	Potiguar	619,9
3-OGX-95-MA	Gavião Branco	Parnaíba	985,4
4-PGN-25-MA	Gavião Tesoura	Parnaíba	1624,9
1-OGX-117-MA	Gavião Tesoura	Parnaíba	1235,3
4-PGN-19-MA	Gavião Tesoura	Parnaíba	5257,2
7-GVB-8-MA	Gavião Branco	Parnaíba	980,1
7-GVB-6-MA	Gavião Branco	Parnaíba	1035,9
7-RUC-59DP-AM	Rio Urucu	Solimões	922,5
3-BRSA-515-AM	Rio Urucu	Solimões	1809,5
3-BRSA-1057DA-AM	Rio Urucu	Solimões	575,1
3-BRSA-670-AM	Rio Urucu	Solimões	1573,1
3-BRSA-1121-AM	Rio Urucu	Solimões	1427,4

### Séries Auto-Organizáveis (SOrS)

As Séries Auto-Organizáveis consistem em uma técnica de auto-organização de séries temporais introduzida por (CARNEIRO, 2021) com base na integração de informações provenientes da posição e da distância obtidos por Mapas Auto-Organizáveis (SOM). Os SOM são notórios pela capacidade que possuem em reduzir dimensionalidade dos dados (HAYKIN, 2008; KOHONEN, 2001). A técnica SOrS se baseia na junção de informações da Matriz-U, obtida a partir da aplicação da técnica SOM, com uma segunda matriz de valores relacionados às posições dos hits da Matriz-U. Essa segunda matriz, denominada de Matriz de Posição (Matriz-P), denota a abordagem inovadora e capaz de possibilitar a indexação da posição topológica toroidal da projeção dos neurônios treinados.

### Ponderações de Posição e Distância para geração de SOrS

Séries baseadas exclusivamente nos índices distância (Índice-D) ou posição (Índice-P), bem como aquelas geradas pela integração entre os índices de posição e distância (IDP), são desenvolvidas com base na posição do BMU de cada uma das instâncias de treinamento, normalizadas para o intervalo de [0,1]. A reconstrução da série temporal, portanto, no Índice-D foi obtido a partir da Matriz-D, construída pela normalização dos valores de dissimilaridade da Matriz-U, enquanto que no Índice-P se baseia na

## 11º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS

normalização da posição de cada BMU no toroide de treinamento, dada pela Matriz-P. Dessa maneira, é gerada uma assinatura baseada no processo de auto-organização denominado SOrS. As séries auto-organizadas resultantes são denominadas, respectivamente, de D-SOrS, P-SOrS e IDP-SOrS.

Após testes com diferentes proporcionalidades, a definição da proporcionalidade de 20% do Índice-D e 80% do Índice-P garantiu ao IDP condições de dar maior ênfase às variedades litoestratigráficas (dadas pela proporção de 80%) em relação às zonas de maior homogeneidade ou heterogeneidade litológica (dadas pela proporção de 20%).

### Similaridade entre Poços de Petróleo por DTW

A similaridade foi performada a partir da combinação par-a-par entre os 30 poços da base de dados de treinamento. A comparação foi aplicada às séries obtidas pela técnica SOrS em duas abordagens, a saber: (i) perfis síntese gerados pela IDP-SOrS; e (ii) combinação vetorial entre as séries D-SOrS e P-SOrS. O algoritmo aplicado para comparar as séries temporais foi o DTW, em razão da sua capacidade de encontrar o alinhamento não linear ótimo entre sequências numéricas que não apresentem necessariamente o mesmo tamanho. O algoritmo de DTW aplicado foi modificado do algoritmo de (KAMPER, 2021) e compilado para aceleração do processamento com base na biblioteca Numba (ANACONDA, 2018).

### Avaliações: análise do custo mínimo de alinhamento por DTW nas ponderações IDP-SOrS e Clusterização Hierárquica

Entende-se que ambas as séries, D-SOrS e P-SOrS, são relevantes na definição de uma série única, IDP-SOrS, que reflete as litologias e as variabilidades litológicas de um poço de petróleo. Para corroborar esta compreensão, foi desenvolvida uma análise de custo mínimo de alinhamento obtido por DTW para cada proporção entre 0 e 100% variando a cada 10% de D-SOrS e P-SOrS. O custo de alinhamento foi calculado para cada uma das 435 comparações possíveis entre os 30 poços de petróleo. Essa análise mostrou que o custo mínimo de alinhamento varia de 0% a 80% de ponderamento para a P-SOrS, denotando, dessa maneira, a relevância da integração ocasionada pela IDP-SOrS.

Com base nas 435 distâncias DTW que representam cada uma das comparações entre as combinações possíveis entre os 30 poços de treinamento, estas distâncias foram organizadas em uma matriz de distâncias [30x30] com valor 0 na diagonal principal. Tal diagonal está relacionada ao custo de alinhamento de uma série temporal consigo mesma. Essa matriz foi utilizada como entrada pré-computada para o cálculo do método aglomerativo hierárquico com a função linkage do sub-módulo hierarchy do módulo cluster do pacote SciPy. Em seguida foi gerado o dendrograma resultante desse cálculo, e escolhido o cut-off capaz de discriminar a quantidade de cinco grupos, relacionados às cinco bacias por estes representados.

## Resultados e Discussão

### Séries Auto-Organizáveis

As SOrS obtidas partiram dos resultados de uma análise SOM. Vale ressaltar, no entanto, que essas séries diferem de uma análise exclusivamente SOM, uma vez que são dotadas de informações relacionadas a uma adaptação da Matriz-U, mas também apresentam informações relacionadas à Matriz-P. A Matriz-IDP (Figura 1) associou as informações de distância fuzzificadas, gerada pela Matriz-D, com as relações de posição, dada pela Matriz-P.

O retorno dos índices gerados pela Matriz-D, Matriz-P e Matriz-IDP ao espaço original, em profundidades, deu origem às séries temporais D-SOrS, P-SOrS e IDP-SOrS. Estas três séries foram performadas para cada um dos 30 poços, e serviram como base para as análises de similaridade sequenciais. A Figura 1 mostra as séries desenvolvidas para a Bacia do Parnaíba.

Figura 1 – (A) Matriz-D, obtida pela fuzzificação da Matriz-U a partir da função sigmoidal, para propiciar maior contraste dos dados de dissimilaridade; (B) Matriz-P, obtida a partir da aplicação do algoritmo Cuthill-McKee em coloração sintética, com periodicidade preservada; (C) Matriz-IDP, obtida a partir da ponderação da Matriz-D com a Matriz-P.

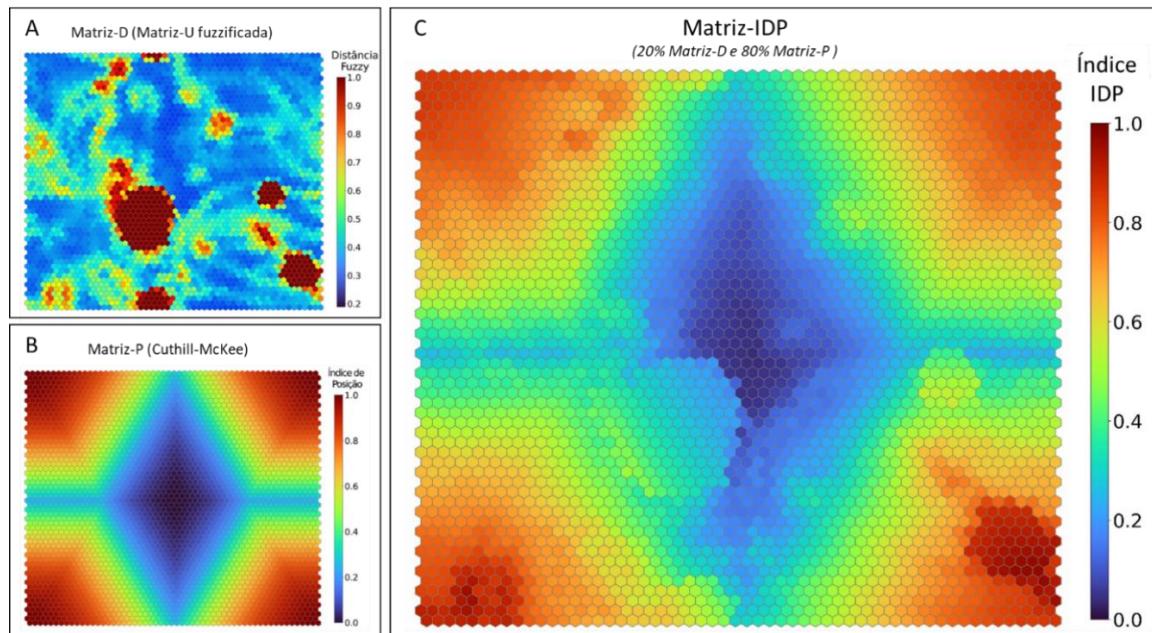
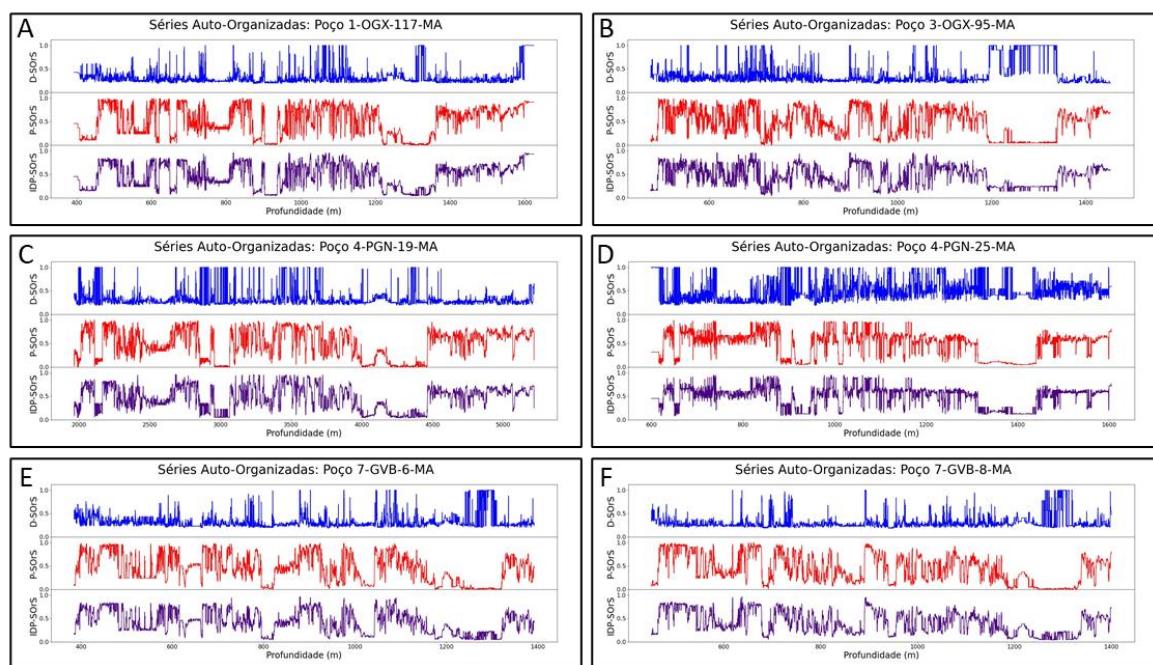


Figura 1 – Séries Auto-Organizadas D-SOrS (azul), P-SOrS (vermelho); e IDP-SOrS (rôxo) desenvolvidas para os poços da Bacia do Parnaíba: (A) 1-OGX-117-MA; (B) 3-OGX-95-MA; (C) 4-PGN-19-MA; (D) 4-PGN-25-MA; (E) 7-GVB-6-MA; e (F) 7-GVB-8-MA.

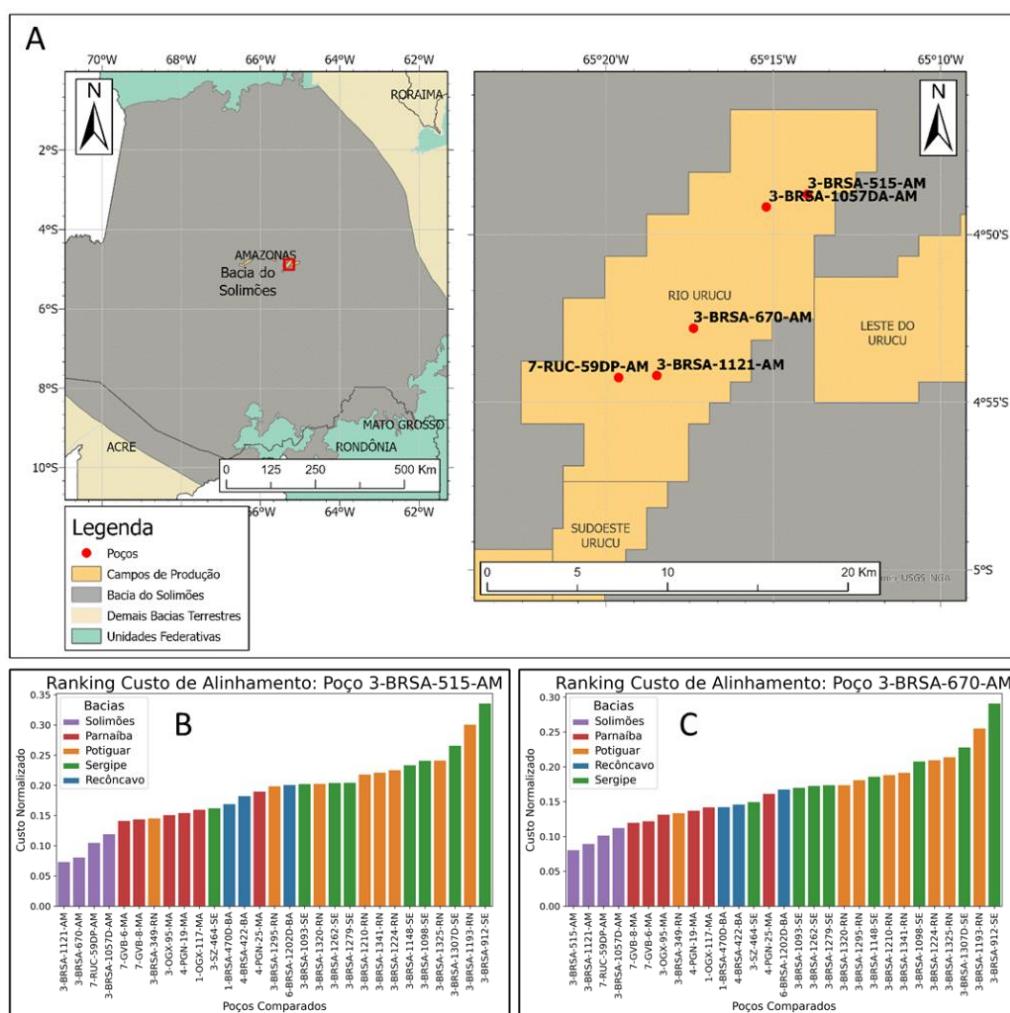


### Rankings de similaridade entre perfis de poços

A análise dos rankings de similaridade entre poços aqui desenvolvida parte do princípio da correlação espacial, que pressupõe que amostras mais próximas entre si no espaço georreferenciado são mais similares do que aquelas que se encontram mais distantes. Vale ressaltar que, nas aplicações de perfilagem de poços, há de se considerar, ainda, a variabilidade do empilhamento litoestratigráfico, o contexto deposicional, além das diferenças de profundidade entre os poços. De forma simplificada, neste trabalho, consideraremos os contextos deposicionais e litoestratigráficos como relacionados às bacias sedimentares e campos de produção relacionados aos poços analisados. É importante ressaltar que bacias localizadas às margens do continente possuem contextos geológicos e empilhamentos estratigráficos parecidos. Portanto, de antemão, é de se esperar que os poços de das bacias Potiguar, do Sergipe e do Recôncavo apresentem contextos litoestratigráficos mais similares entre si do que as bacias intracontinentais do Solimões e Parnaíba. Esse entendimento deverá ser considerado na análise dos rankings desenvolvida a seguir.

Os poços da Bacia do Solimões apresentaram baixo custo normalizado de alinhamento DTW entre si. Todos estes poços estão em um mesmo campo de produção (Rio Urucu), conforme a Figura 3A. O primeiro poço analisado, 3-BRSA-515-AM (1809,5 m), Figura 3B, apresenta maior similaridade, respectivamente, aos poços 3-BRSA-1121-AM (1427,4 m) e 3-BRSA-670-AM (1573,1 m).

Figura 2 – (A) Disposição espacial dos poços da Bacia do Solimões; Rankings de similaridade baseado no custo de alinhamento para os poços: (B) 3-BRSA-515-AM; (C) 3-BRSA-670-AM.



## **Conclusões**

A sistemática de criação das SOrS, propostas nesta pesquisa, permitiram a análise de perfis de poços, aqui tratados como séries temporais multivariadas, com base na auto-organização SOM. A geração da D-SOrS, P-SOrS e IDP-SOrS representando, respectivamente, índices de distância, posição, e uma ponderação entre distância e posição, propiciou a consequente reconstrução dessas séries auto-organizadas no domínio das profundidades originais relacionadas aos perfis. A partir destas séries auto-organizadas foi possível observar as similaridades entre um sinal representativo de todas as variáveis de treinamento. Contudo, essa comparação é fortemente dependente da capacidade do algoritmo SOM discretizar, no mapa auto-organizado, os dados de entrada. À vista disso a avaliação dos impactos dos parâmetros de treinamento SOM na assinatura unificada, como o tamanho do mapa de treinamento, ainda necessitam ser melhor explorados.

As SOrS viabilizaram assinaturas unificadas e comparáveis, com resultados coerentes e satisfatórios para os 30 poços selecionados analisados. As séries geradas preservaram a similaridade litoestratigráfica observada nestes poços. Além disso, as séries foram capazes de discretizar grupos de similaridade com significado geológico, conforme a medida do custo de alinhamento provida pelo algoritmo DTW. Vale ressaltar que a interpretabilidade das SOrS é fortemente dependente das variáveis escolhidas para o treinamento, uma vez que diferentes variáveis geram relações diferentes de posição e distância no processo de auto-organização.

## **Referências Bibliográficas**

- ALI, M. et al. Machine learning - A novel approach of well logs similarity based on synchronization measures to predict shear sonic logs. *Journal of Petroleum Science and Engineering*, v. 203, p. 108602, ago. 2021.
- ANACONDA. Numba: A High Performance Python Compiler. Disponível em: <<https://numba.pydata.org/>>. 2018.
- CARNEIRO, C. C. Ciência de dados para determinação da similaridade entre perfis de poços de petróleo: uma abordagem a partir de Mapas Auto-Organizáveis. Trabalho de conclusão de curso (MBA em Ciência de Dados). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. 2021.
- DU, R. et al. A similarity measure recognized by morphological characteristics analysis of well logging curves: application to the knowledge domain of sandstone reservoir. *Arabian Journal of Geosciences*, v. 13, n. 18, p. 947, set. 2020.
- HAYKIN, S. *Neural Networks and Learning Machines*. 3a Edição ed. New York: Prentice Hall, 2008.
- KAMPER, H. *Dynamic Time Warping Application*, 2021.
- KOHONEN, T. *Self-Organizing Maps*. 3. ed. Berlin Heidelberg: Springer-Verlag, 2001.